Algebraic recognizability of regular tree languages

Zoltán Ésik [a,1] Pascal Weil [b,2]

[a] *Department of Computer Science, University of Szeged, Hungary Research Group on Mathematical Linguistics, Rovira i Virgili University, Tarragona, Spain.*

[b] *LaBRI, CNRS, Université Bordeaux-1, France*

Abstract

We propose a new algebraic framework to discuss and classify recognizable tree languages, and to characterize interesting classes of such languages. Our algebraic tool, called preclones, encompasses the classical notion of syntactic Σ-algebra or minimal tree automaton, but adds new expressivity to it. The main result in this paper is a variety theorem *à la* Eilenberg, but we also discuss important examples of logically defined classes of recognizable tree languages, whose characterization and decidability was established in recent papers (by Benedikt and Ségoufin, and by Bojańczyk and Walukiewicz) and can be naturally formulated in terms of pseudovarieties of preclones. Finally, this paper constitutes the foundation for another paper by the same authors, where first-order definable tree languages receive an algebraic characterization.

1 Introduction

The notion of recognizability emerged in the 1960s (Eilenberg, Mezei, Wright, and others, cf. [17,30]) and has been the subject of considerable attention since,

Email addresses: `ze@inf.u-szeged.hu` (Zoltán Ésik), `pascal.weil@labri.fr` (Pascal Weil).

[1] Partial support from the National Foundation of Hungary for Scientific Research, grant T46686 is gratefully acknowledged.

[2] Partial support from the *ACI Sécurité Informatique* (projet VERSYDIS) of the French Ministère de la Recherche is gratefully acknowledged. Part of this work was done while P. Weil was an invited professor at the University of Nebraska in Lincoln.

notably because of its close connections with automata-theoretic formalisms and with logical definability, cf. [6,15,18,38] for some early papers.

Recognizability was first considered for sets (languages) of finite words, cf. [16] and the references contained in op. cit. The general idea is to use the algebraic structure of the domain, say, the monoid structure on the set of all finite words, to describe some of its subsets, and to use algebraic considerations to discuss the combinatorial or logical properties of these subsets. More precisely, a set of words is said to be recognizable if it is a union of classes in a (locally) finite congruence. The same concept was adapted to the case of finite trees, traces, finite graphs, etc, cf. [17,30,14,9], where it always entertains close connections with logical definability [11,12].

It follows rather directly from this definition of (algebraic) recognizability that a finite – or finitary – algebraic structure can be canonically associated with each recognizable subset L, called its syntactic structure. Moreover, the algebraic properties of the syntactic structure of L reflect its combinatorial and logical properties. The archetypal example is that of star-free languages of finite words: they are exactly the languages whose syntactic monoid is aperiodic, cf. [34]. They are also exactly the languages that can be defined by a first-order sentence of the predicate $<$ ($FO[<]$), cf. [29], and the languages that can be defined by a temporal logic formula, cf. [27,22,7]. In particular, every algorithm we know for deciding the $FO[<]$-definability of a regular language L, works by checking, more or less explicitly, whether the syntactic monoid of L is aperiodic.

Let Σ be a ranked alphabet. In this paper, we are interested in sets of finite Σ-labeled trees, or tree languages. It has been known since the 1960s [17,30,15] that the tree languages that are definable in monadic second order logic are exactly the so-called regular tree languages, that is, those accepted by bottom-up tree automata. Moreover, deterministic tree automata suffice to accept these languages, and each regular tree language admits a unique, minimal deterministic automaton. From the algebraic point of view, the set of all Σ-labeled trees can be viewed in a natural way as a (free) Σ-algebra, where Σ is now seen as a signature. Moreover, a deterministic bottom-up tree automaton can be identified with a finite Σ-algebra, with some distinguished (final) elements. Thus regular tree languages are also the recognizable subsets of the free Σ-algebra.

The situation however is not entirely satisfying, because we know very little about the structure of finite Σ-algebras, and very few classes of tree languages have been characterized in algebraic terms, see [26,32,33] for attempts to use Σ-algebra-theoretic considerations (and some variants) for the purpose of classifying tree languages. In particular, the important problem of deciding whether a regular tree language is $FO[<]$-definable remained open [33]. Based

on the word language case, it is tempting to guess that an answer to this problem ought to be found using algebraic methods.

In this paper, we introduce a new algebraic framework to handle tree languages. More precisely, we consider algebras called preclones (they lack some of the operations and axioms of clones [13]). Precise definitions are given in Section 2.1. Let us simply say here that, in contrast with the more classical monoids or Σ-algebras, preclones have infinitely many sorts, one for each integer $n \geq 0$. As a result, there is no nontrivial finite preclone. The corresponding notion is that of finitary preclones, that have a finite number of elements of each sort. An important class of preclones is given by the transformations $T(Q)$ of a set Q. The elements of sort (or rank) n are the mappings from Q^n into Q and the (preclone) composition operation is the usual composition of mappings. Note that $T(Q)$ is finitary if Q is finite.

It turns out that the finite Σ-labeled trees can be identified with the 0-sort of the free preclone generated by Σ. The naturally defined syntactic preclone of a tree language L is finitary if and only if L is regular. In fact, if S is the syntactic Σ-algebra of L, the syntactic preclone is the sub-preclone of $T(S)$ generated by the elements of Σ (if $\sigma \in \Sigma$ is an operation of rank r, it defines a mapping from S^r into S, and hence an element of sort r in $T(S)$). Note that this provides an effectively constructible description of the syntactic preclone of L.

It is important to note that the class of recognizable tree languages in the preclone-theoretic sense, is exactly the same as the usual one – we are simply adding more algebraic structure to the finitary minimal object associated with a regular tree language, and thus, we give ourselves a more expressive language to capture families of tree languages.

In order to justify the introduction of such an algebraic framework, we must show not only that it offers a well-structured framework, that accounts for the basic notions concerning tree languages, but also that it allows the characterization of interesting classes of tree languages. The first objective is captured in the definition of varieties of tree languages, and their connection with pseudovarieties of finitary preclones, by means of an Eilenberg-type theorem. This is not unexpected, but it requires combinatorially much more complex proofs than in the classical word case, the details of which can be found below in Section 5.1.

As for the second objective, we offer several elements. First the readers will find in this paper a few simple but hopefully illuminating examples, which illustrate similarities and differences with the classical examples from the theory of word languages. Second, we discuss a couple of important recent results on the characterization of certain classes of tree languages: one concerns the tree lan-

guages that are definable in the first-order logic of successors ($FO(\mathbf{Succ})$), and is due to Benedikt and Ségoufin [3]; the second one concerns the tree languages defined in the logics EF and EX, and is due to Bojańczyk and Walukiewicz [5]. Neither of these remarkable results can be expressed directly in terms of syntactic Σ-algebras; neither mentions preclones (of course) but both use mappings of arity greater than 1 on Σ-algebras, that is, they can be naturally expressed in terms of preclones, as we explain in Sections 5.2.2 and 5.2.3. It is also very interesting to note that the conditions that characterize these various classes of tree languages include the semigroup-theoretic characterization of their word language analogues, but cannot be reduced to them.

Another such result, and that was our original motivation to introduce the formalism of preclones, is a nice algebraic characterization of $FO[<]$-definable tree languages (and a number of extensions of $FO[<]$, such as the introduction of additional, modular quantifiers), briefly discussed in Section 5.2.4. Let us say immediately that we do not know yet whether this characterization can be turned into a decision algorithm! In order to keep this paper within a reasonable number of pages, this characterization will be the subject of another paper by the same authors [21]. The main results of this upcoming paper can be found, along with an outline of the present paper, in [20].

To summarize the plan of the paper, Section 2 introduces the algebraic framework of preclones, discussing in particular the all-important cases of free preclones, in which tree languages live (Section 2.2), and of preclones associated with tree automata (Section 2.3). Section 2.4 discusses in some details the notion of finite determination for a preclone, a finiteness condition different from being finitary, which is crucial in the sequel. Section 2.5 is included for completeness (and can be skipped at first reading): its aim is to make explicit the connection between our preclones and other known algebraic structures, namely magmoids and strict monoidal categories.

Recognizable tree languages are the subject of Section 3. Here tree languages are meant to be any subset of some ΣM_k, and the preclone structure on ΣM naturally induces a notion of recognizability, as well as a notion of syntactic preclone (Section 3.1). As pointed out earlier, the usual recognizable tree languages, that is, subsets of ΣM_0, fall nicely in this framework, and there is a tight connection between the minimal automaton of such a language and its syntactic preclone (Section 3.2). Specific examples are given in Section 3.3.

Pseudovarieties of finitary preclones are discussed in detail in Section 4. As it turns out, this notion is not a direct translate of the classical notion for semigroups or monoids, due to the infinite number of sorts. The technical treatment of these classes is rather complex, and we deal with it thoroughly, since it is the foundation of our construction. We show in particular that pseudovarieties are characterized by their finitely determined elements (Section 4.1),

and we describe the pseudovarieties generated by a given set of finitary preclones, showing in particular that membership in a 1-generated pseudovariety is decidable (Section 4.2).

Finally, we introduce varieties of tree languages and we establish the variety theorem in Section 5.1. Section 5.2 presents the examples described above, based on the results by Benedikt and Ségoufin [3] and by Bojańczyk and Walukiewicz [5].

2 The algebraic framework

In this section, we introduce the notion of preclones, a multi-sorted kind of algebra which is our central tool in this paper. In the sequel, if n is an integer, $[n]$ denotes the set of integers $\{1, \ldots, n\}$. In particular, $[0]$ denotes the empty set.

2.1 Preclones and preclone-generators pairs

Let Q be a set and let $T_n(Q)$ denote the set of n-ary transformations of Q, that is, mappings from Q^n to Q. Let then $T(Q)$ be the sequence of sets of transformations $T(Q) = (T_n(Q))_{n\geq 0}$, which will be called the *preclone of transformations* of Q. The set $T_1(Q)$ of transformations of Q is a monoid under the composition of functions. Composition can be considered on $T(Q)$ in general: if $f \in T_n(Q)$ and $g_i \in T_{m_i}(Q)$ $(1 \leq i \leq n)$, then the composite $h = f(g_1, \ldots, g_n)$, defined in the natural way, is an element of $T_m(Q)$ where $m = \sum_{i\in[n]} m_i$:

$$h(q_{1,1}, \ldots, q_{n,m_n}) = f(g_1(q_{1,1}, \ldots, q_{1,m_1}), \ldots, g_n(q_{n,1}, \ldots, q_{n,m_n}))$$

for all $q_{i,j} \in Q$, $1 \leq i \leq n$, $1 \leq j \leq m_i$. This composition operation and its associativity properties are exactly what is captured in the notion of a preclone.

In general, a *preclone* is a many-sorted algebra $S = ((S_n)_{n\geq 0}, \bullet, \mathbf{1})$. The elements of the sets S_n, where n ranges over the nonnegative integers, are said to have *rank* n. The *composition operation* $\bullet$ associates with each $f \in S_n$ and $g_1 \in S_{m_1}, \ldots, g_n \in S_{m_n}$, an element $\bullet(f, g_1, \ldots, g_n) \in S_m$, of rank $m = \sum_{i\in[n]} m_i$. We usually write $f \cdot (g_1 \oplus \cdots \oplus g_n)$ for $\bullet(f, g_1, \ldots, g_n)$. Finally, the constant $\mathbf{1}$ is in S_1. Moreover, we require the following three equational axioms:

$$(f \cdot (g_1 \oplus \cdots \oplus g_n)) \cdot (h_1 \oplus \cdots \oplus h_m) = f \cdot ((g_1 \cdot \overline{h}_1) \oplus \cdots \oplus (g_n \cdot \overline{h}_n)), \quad (1)$$

where $f, g_1, \ldots, g_n$ are as above, $h_j \in S_{k_j}$ ($j \in [m]$), and if we denote $\sum_{j \in [i]} m_j$ by $m_{[i]}$, then $\overline{h}_i = h_{m_{[i-1]}+1} \oplus \cdots \oplus h_{m_{[i]}}$ for each $i \in [n]$;

$$\mathbf{1} \cdot f = f \quad (2)$$

$$f \cdot (\mathbf{1} \oplus \cdots \oplus \mathbf{1}) = f, \quad (3)$$

where $f \in S_n$ and $\mathbf{1}$ appears n times on the left hand side of the last equation.

Note that Axiom (1) generalizes associativity, and Axioms (2) and (3) can be said to state that $\mathbf{1}$ is a neutral element.

Remark 2.1 The elements of rank 1 of a preclone form a monoid. □

It is immediately verified that $T(Q)$, the preclone of transformation of a set Q, is indeed a preclone for the natural composition of functions, with the identity function id_Q as $\mathbf{1}$. Preclones are an abstraction of sets of n-ary transformations of a set, which generalizes the abstraction from transformation monoids to monoids.

Remark 2.2 Clones [13], or equivalently, Lawvere theories [4,19] are another more classical abstraction. Readers interested in the comparison between clones and preclones will have no difficulty tracing their differences in the sequel. We will simply point out here the fact that, in contrast with the definition of the clone of transformations of Q, each of the m arguments of the composite $f(g_1, \ldots, g_n)$ above is used in exactly one of the g_i's, the first m_1 in g_1, the next m_2 in g_2, etc. □

We observe that a preclone with at least one element of rank greater than 1 must have elements of arbitrarily high rank, and hence cannot be finite. We say that a preclone S is *finitary* if and only if each S_n is finite. For instance, the preclone of transformations $T(Q)$ is finitary if and only if the set Q is finite.

The notions of *morphism* between preclones, *sub-preclone*, *congruence* and *quotient* are defined as usual [25,39]. Note that, as is customary for multi-sorted algebras, a morphism maps elements of rank n to elements of the same rank, and a congruence only relates elements of the same rank.

To facilitate discussions, we introduce the following short-hand notation. An n-tuple $(g_1, \ldots, g_n)$ of elements of S will often be written as a formal $\oplus$-sum: $g_1 \oplus \cdots \oplus g_n$. Moreover, if $g_i \in S_{m_i}$ ($1 \leq i \leq n$), we say that $g_1 \oplus \cdots \oplus g_n$ has *total rank* $m = \sum_{i \in [n]} m_i$. Finally, we denote by $S_{n,m}$ the set of all n-tuples of

total rank m. With this notation, $S_{1,n} = S_n$. The n-tuple $\mathbf{1} \oplus \cdots \oplus \mathbf{1} \in S_{n,n}$ is denoted by $\mathbf{n}$. If G is a subset of S, we also denote by $G_{n,m}$ the set of n-tuples of elements of G, of total rank m.

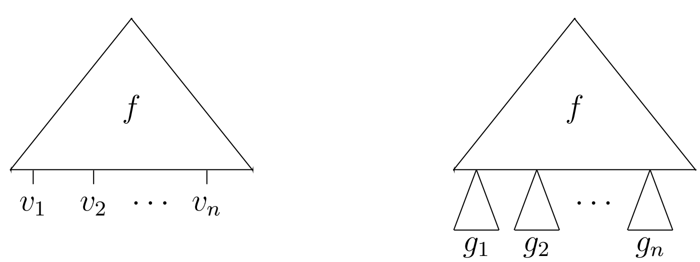
Observe that a preclone morphism $\varphi: S \to T$, naturally extends to a map $\varphi: S_{n,m} \to T_{n,m}$ for each $n, m \geq 0$, by mapping $g_1 \oplus \cdots \oplus g_n$ to $\varphi(g_1) \oplus \cdots \oplus \varphi(g_n)$.

For technical reasons, it will often be preferable to work with pairs (S, A) consisting of a preclone S and a (possibly empty) set A of generators of S. We call such pairs *preclone-generators pairs*, or *pg*-pairs. The notions of morphisms and congruences must be revised accordingly: in particular, a morphism of *pg*-pairs from (S, A) to (T, B) must map A into B. A *pg*-pair (S, A) is said to be finitary if S is finitary and A is finite.

Besides preclones of transformations of the form $T_n(Q)$, fundamental examples of preclones and *pg*-pairs are the free preclones and the preclones associated with a tree automaton. These are discussed in the next sections.

2.2 Trees and free preclones

Let Σ be a ranked alphabet, say, $\Sigma = (\Sigma_n)_{n \geq 0}$, and let $(v_k)_{k \geq 1}$ be a sequence of variable names. We let ΣM_n be the set of finite trees whose inner nodes are labeled by elements of Σ (according to their rank), whose leaves are labeled by elements of $\Sigma_0 \cup \{v_1, \ldots, v_n\}$, and whose *frontier* (the left to right sequence of variables appearing in the tree) is the word $v_1 \cdots v_n$: that is, each variable occurs exactly once, and in the natural order. Note that ΣM_0 is the set of finite Σ-labeled trees. We let $\Sigma M = (\Sigma M_n)_n$.



If $f \in \Sigma M_n$ and $g_1, \ldots, g_n \in \Sigma M$, the composite tree $f \cdot (g_1 \oplus \cdots \oplus g_n)$ is obtained by substituting the root of the tree g_i for the variable v_i in f for each i, and renumbering consecutively the variables in the frontiers of $g_1, \ldots, g_n$. Let also $\mathbf{1} \in \Sigma M_1$ be the tree with a single vertex, labeled v_1. Then $(\Sigma M, \cdot, \mathbf{1})$ is a preclone.

Each letter $\sigma \in \Sigma$ of rank n can be identified with the tree with root labeled

σ, where the root's children are leaves labeled $v_1, \ldots, v_n$. It is easily verified that every rank-preserving map from Σ to a preclone S can be extended in a unique fashion to a preclone morphism from ΣM into S. That is:

Proposition 2.3 *ΣM is the free preclone generated by Σ, and $(\Sigma M, \Sigma)$ is the free pg-pair generated by Σ.*

Remark 2.4 If $\Sigma_n = \emptyset$ for each $n \neq 1$, then $\Sigma M_n = \emptyset$ for all $n \neq 1$, and ΣM_1 can be assimilated with the set of all finite words on the alphabet Σ_1.

If at least one Σ_n with $n > 1$ is nonempty, then infinitely many ΣM_n are nonempty, and if in addition $\Sigma_0 \neq \emptyset$, then each ΣM_n is nonempty. □

2.3 Examples of preclones

We already discussed preclones of transformations and free preclones. The next important class of examples is that of preclones (and *pg*-pairs) associated with Σ-algebras and tree automata. We also discuss a few simple examples of preclones that will be useful in the sequel.

2.3.1 Preclone associated with an automaton

Let Σ be a ranked alphabet as above and let Q be a Σ-*algebra*: that is, Q is a set and each element $\sigma \in \Sigma_n$ defines an n-ary transformation of Q, *i.e.*, a mapping $\sigma^Q \colon Q^n \to Q$. Recall that Q, equipped with a set $F \subseteq Q$ of *final states*, can also be described as a (deterministic, bottom-up) *tree automaton* accepting trees in ΣM_0, cf. [15,38,23,24,8].

More precisely, the mapping $\sigma \mapsto \sigma^Q$ induces a morphism of Σ-algebras from ΣM_0, viewed here as the initial Σ-algebra (*i.e.*, the algebra of Σ-terms), to Q, say, $\mathsf{val} \colon \Sigma M_0 \to Q$, and the tree language accepted by Q is the set $\mathsf{val}^{-1}(F)$ of trees which evaluate to an element of F.

Now, since the elements of Σ_n can be viewed also as elements of $T_n(Q)$, the mapping $\sigma \mapsto \sigma^Q$ also extends to a preclone morphism $\tau \colon \Sigma M \to T(Q)$, whose restriction to the rank 0 elements is exactly the morphism val. The range of τ is called the *preclone associated with* Q, and the *pg-pair associated with* Q, written $\mathsf{pg}(Q)$, is the pair $(\tau(\Sigma M), \tau(\Sigma))$.

We observe in particular that a morphism of Σ-algebras $\varphi \colon Q \to Q'$ induces a morphism of *pg*-pairs $\varphi \colon \mathsf{pg}(Q) \to \mathsf{pg}(Q')$ in a functorial way.

Conversely, if Q is a set and $\tau \colon \Sigma M \to T(Q)$ is a preclone morphism such

that $\tau(\Sigma M_0) = Q$, letting $\sigma^Q = \tau(\sigma)$ endows the set Q with a structure of Σ-algebra, for which the associated preclone is the range of τ.

In the sequel, when discussing decidability issues concerning preclones, we will say that a preclone is *effectively given* if it is given as the preclone associated with a finite Σ-algebra Q, that is, by a finite set of generators in $T(Q)$. By definition, such a preclone is finitary.

2.3.2 Simple examples of preclones

The following examples of preclones and *pg*-pairs will be discussed throughout the rest of this paper.

Example 2.5 Let $\mathbb{B}$ be the 2-element set $\mathbb{B} = \{\mathsf{true}, \mathsf{false}\}$, and let $T_\exists$ be the subset of $T(\mathbb{B})$ whose rank n elements are the n-ary or function and the n-ary constant true, written respectively or_n and true_n. One verifies easily that $T_\exists$ is a preclone, which is generated by the binary or_2 function and the nullary constants true_0 and false_0. That is, if Σ consists of these three generators, $T_\exists$ is the preclone associated with the Σ-automaton whose state set is $\mathbb{B}$.

Moreover, the rank 1 elements of $T_\exists$ form a 2-element monoid, isomorphic to the multiplicative monoid $\{0, 1\}$, and known as U_1 in the literature on monoid theory, e.g. [31]. $\square$

Example 2.6 Let p be an integer, $p \geq 2$ and let $\mathbb{B}_p = \{0, 1, \ldots, p-1\}$. We let T_p be the subset of $T(\mathbb{B}_p)$ whose rank n elements ($n \geq 0$) are the mappings $f_{n,r}\colon (r_1, \ldots, r_n) \mapsto r_1 + \cdots + r_n + r \bmod p$ for $0 \leq r < p$. It is not difficult to verify that T_p is a preclone, and that it is generated by the nullary constant 0, the unary increment function $f_{1,1}$ and the binary sum $f_{2,0}$.

As in Example 2.5, T_p can be seen as the preclone associated with a p-state automaton. Moreover, its rank 1 elements form a monoid isomorphic to the cyclic group of order p. $\square$

Example 2.7 Let again $\mathbb{B} = \{\mathsf{true}, \mathsf{false}\}$, and let T_{path} be the subset of $T(\mathbb{B})$ whose rank 0 elements are the nullary constants true_0 and false_0, and the rank n elements ($n > 0$) are the n-ary constants true_n and false_n, and the n-ary partial disjunctions or_P (if $P \subseteq [n]$, or_P is the disjunction of the i-th arguments, $i \in P$). One verifies easily that T_{path} is a preclone, which is generated by the binary or_2 function, the nullary constants true_0 and false_0, and the unary constant false_1. The rank 1 elements of T_{path} form a 3-element monoid, isomorphic to the multiplicative monoid $\{1, a, b\}$ with $xy = y$ for $x, y \neq 1$, known as U_2 in the literature on monoid theory, e.g. [31]. $\square$

Section 2.3.1 shows the importance of the representation of preclones as preclones of transformations. It is not difficult to establish the following analogue of Cayley's theorem.

Proposition 2.8 *Every preclone can be embedded in a preclone of transformations.*

Proof. Suppose that S is a preclone and let Q be the disjoint union of the sets S_n, $n \geq 0$. For each $f \in S_n$, let $\overline{f}$ be the function $Q^n \to Q$ given by

$$\overline{f}(g_1, \ldots, g_n) = f \cdot (g_1 \oplus \cdots \oplus g_n).$$

The assignment $f \mapsto \overline{f}$ defines an injective morphism $S \to T(Q)$. $\square$

This result however is not very satisfactory: it does not tell us whether a finitary preclone can be embedded in the preclone of transformations of a finite set. It is actually not always the case, and this leads to the following discussion.

Let $k \geq 0$. We say that a preclone S is *k-determined* if distinct elements can be separated by k-ary equations. Formally, let $\sim_k$ denote the following equivalence relation: for all $f, g \in S_n$ $(n \geq 0)$,

$$f \sim_k g \Longleftrightarrow f \cdot h = g \cdot h, \text{ for all } h \in S_{n,\ell} \text{ with } \ell \leq k.$$

Note that for each $\ell \leq k$, $\sim_k$ is the identity relation on T_ℓ. We call S *k-determined* if the relation $\sim_k$ is the identity relation on each S_n, $n \geq 0$, and we say that S is *finitely determined* if it is k-determined for some integer k. We also say that a *pg*-pair (S, A) is k-determined (resp. finitely determined) if S is.

Example 2.9 The preclone of transformations of a set is 0-determined. $\square$

We observe the two following easy lemmas.

Lemma 2.10 *For each k, $\sim_k$ is a congruence relation.*

Proof. Let $f, g \in S_n$ be $\sim_k$-equivalent. For each $i \in [n]$, let $f_i, g_i \in S_{m_i}$, such that $f_i \sim_k g_i$. We want to show that $f \cdot (f_1 \oplus \cdots \oplus f_n) \sim_k g \cdot (g_1 \oplus \cdots \oplus g_n)$.

Let $m = \sum_{i \in [n]} m_i$ and let $h \in S_{m,\ell}$ for some $\ell \leq k$. Then h is an m-tuple, and we let h_1 be the tuple of the first m_1 terms of h, h_2 consist of the next m_2

elements, etc, until h_n, which consists of the last m_n elements of h. Note that each h_i lies in some S_{m_i,ℓ_i} and that $\sum_{i\in[n]} \ell_i = \ell$. In particular, $\ell_i \leq k$ for each i and we have

$$\begin{aligned} f \cdot (f_1 \oplus \cdots \oplus f_n) \cdot h &= f \cdot (f_1 \cdot h_1 \oplus \cdots \oplus f_n \cdot h_n) \\ &= f \cdot (g_1 \cdot h_1 \oplus \cdots \oplus g_n \cdot h_n) \\ &= g \cdot (g_1 \cdot h_1 \oplus \cdots \oplus g_n \cdot h_n) \\ &= g \cdot (g_1 \oplus \cdots \oplus g_n) \cdot h. \end{aligned}$$

□

Lemma 2.11 *For each $k \geq 0$, the quotient preclone $S/{\sim_k}$ is k-determined.*

Proof. Let $T = S/{\sim_k}$ and let $[f], [g] \in T_n$, where $[f]$ denotes the $\sim_k$-equivalence class of f (necessarily in S_n). Let $\ell \leq k$ and assume that $[f] \cdot [h] = [g] \cdot [h]$ for each $h \in T_{n,\ell}$. Then $f \cdot h \sim_k g \cdot h$ for each h. But $f \cdot h$ and $g \cdot h$ lie in S_ℓ, and we already noted that $\sim_k$ is the identity relation on S_ℓ (since $\ell \leq k$). Thus $f \cdot h = g \cdot h$ for all $h \in S_{n,\ell}$, and since this holds for each $\ell \leq k$, we have $f \sim_k g$, and hence $[f] = [g]$. □

We say that a preclone morphism $\varphi\colon S \to T$ is *k-injective* if it is injective on each S_ℓ with $\ell \leq k$. The next lemma, relating k-determination and k-injectivity, will be used to discuss embeddability of a finitary preclone in the preclone of transformations of a finite set.

Lemma 2.12 *Let S be a k-determined preclone. If $\varphi\colon S \to T$ is a k-injective morphism, then φ is injective.*

Proof. If $\varphi(f) = \varphi(g)$ for some $f, g \in S_n$, then $\varphi(f \cdot h) = \varphi(f) \cdot \varphi(h) = h(g) \cdot \varphi(h) = \varphi(g \cdot h)$, for all $h \in S_{n,\ell}$ with $\ell \leq k$. Since φ is k-injective, it follows $f \cdot h = g \cdot h$ for all $h \in S_{n,\ell}$ with $\ell \leq k$, and since S is k-determined, this implies $f = g$. □

Proposition 2.13 *Let S be a finitary and finitely determined preclone. Then there is a finite set Q such that S embeds in $T(Q)$. If in addition S is 0-determined, the set Q can be taken equal to S_0.*

Proof. We modify the construction in the proof of Proposition 2.8. Let $k \geq 0$ be such that S is k-determined, and let $Q = \{\bot\} \cup \bigcup_{i\leq k} S_i$, where the sets S_i are assumed to be pairwise disjoint and $\bot$ is a new symbol, not in any of those sets. For each $f \in S_n$ ($n \geq 0$), let $\varphi(f) = \overline{f} : Q^n \to Q$ be the function

defined by

$$\overline{f}(q_1,\ldots,q_n) = \begin{cases} f\cdot(q_1\oplus\cdots\oplus q_n) & \text{if } q_1\in S_{m_1},\ldots,q_n\in S_{m_n} \text{ and } \sum m_i\le k,\\ \bot & \text{otherwise.}\end{cases}$$

It is easy to check that φ is a morphism. By Lemma 2.12, φ is injective.

To conclude, we observe that if $k=0$, we can choose $Q=S_0$ since $\sum_{i=1}^n m_i\le 0$ is possible if and only if each $m_i=0$. □

For later use we also note the following technical results.

Proposition 2.14 *Let S and T be preclones, with T k-determined. Let G be a (ranked) generating set of S and let $\varphi\colon G\to T$ be a rank-preserving map, whose range includes all of T_ℓ, for each $\ell\le k$. Then φ can be extended to a preclone morphism $\overline{\varphi}\colon S\to T$ iff for all $g\in G_n$, $n\ge 0$, and for all $h\in G_{n,\ell}$ with $\ell\le k$,*

$$\varphi(g\cdot h)=\varphi(g)\cdot\varphi(h). \tag{4}$$

Proof. Condition (4) is obviously necessary, and we show that it is sufficient.

Let $f\in S_n$, $n\ge 0$. Any possible image of f by a preclone morphism is an element $g\in T_n$ such that, if $h\in T_{n,\ell}$ for some $\ell\le k$ and if $h'\in G_{n,\ell}$ is such that $\varphi(h')=h$, then $g\cdot h=\varphi(f\cdot h')$. Since T is k-determined and each T_ℓ ($\ell\le k$) is in the range of φ, the element g is completely determined by f. That is, if an extension of φ exists, then it is unique.

We now show the existence of this extension. We want to assign an image to an arbitrary element f of S, and we proceed by induction on the height of an expression of f in terms of the elements of G; such an expression exists since G generates S. If $f\in G$, we let $\overline{\varphi}(f)=\varphi(f)$. Note also that if $f=\mathbf{1}$, then we let $\overline{\varphi}(f)=\mathbf{1}$. If $f\notin G$, then $f=g\cdot h$ for some $g\in G\cap S_n$ and some $h=h_1\oplus\cdots\oplus h_n$. By induction, the elements $\overline{\varphi}(h_i)$ are well defined for each $i\in[n]$. We then let $\overline{\varphi}(f)=\varphi(g)\cdot(\overline{\varphi}(h_1)\oplus\cdots\oplus\overline{\varphi}(h_n))$.

To show that $\overline{\varphi}(f)$ is well-defined, we consider a different decomposition of f, say, $f=g'\cdot h'$ with $h'=h'_1\oplus\cdots\oplus h'_m$. If f has rank $\ell\le k$, then $h\in S_{n,\ell}$ and $h'\in S_{m,\ell}$, so $h=\varphi(\bar h)$ and $h'=\varphi(\bar h')$ for some $\bar h\in G_{n,\ell}$ and $\bar h'\in G_{m,\ell}$. By Condition (4), we have

$$\varphi(g)\cdot\overline{\varphi}(\bar h)=\varphi(g)\cdot\varphi(\bar h)=\varphi(g\cdot h)=\varphi(f),$$

and by symmetry, $\varphi(g)\cdot\overline{\varphi}(\bar h)=\varphi(g')\cdot\overline{\varphi}(\bar h')$. So $\overline{\varphi}$ is well defined on all the elements of S of rank at most k.

Now if f has rank $\ell > k$, let $x \in T_{\ell,p}$ with $p \leq k$. Then there exists $x' \in G_{\ell,p}$ such that $x = \varphi(x')$. Note that $h \cdot x'$ and $h' \cdot x'$ are well defined, in $S_{n,p}$ and in $S_{m,p}$ respectively. In particular, $\overline{\varphi}(h \cdot x')$ is well defined, and equal to $\overline{\varphi}(h) \cdot x$. Similarly, $\overline{\varphi}(h' \cdot x')$ is well defined, equal to $\overline{\varphi}(h') \cdot x$. It follows that

$$(\varphi(g) \cdot \overline{\varphi}(h)) \cdot x = \varphi(g) \cdot (\overline{\varphi}(h) \cdot x) = \varphi(g) \cdot \overline{\varphi}(h \cdot x') = \overline{\varphi}(g \cdot (h \cdot x')) = \overline{\varphi}(f \cdot x').$$

By symmetry, $(\varphi(g') \cdot \overline{\varphi}(h')) \cdot x = (\varphi(g) \cdot \overline{\varphi}(h)) \cdot x$, and since T is k-determined, we have $\varphi(g) \cdot \overline{\varphi}(h) = \varphi(g') \cdot \overline{\varphi}(h')$. Thus, $\overline{\varphi}$ is well defined on S.

By essentially the same argument, one verifies that $\overline{\varphi}$ preserves composition. □

Corollary 2.15 *Let S and T be k-determined preclones that are generated by their elements of rank at most k. If there exist bijections from S_ℓ to T_ℓ for each $\ell \leq k$, that preserve all compositions of the form $f \cdot g$, where $f \in S_n$, $g \in S_{n,\ell}$ with $n, \ell \leq k$, then S and T are isomorphic.*

Proof. Using the fact that S is generated by its elements of rank at most k and T is k-determined, Proposition 2.14 shows that the given bijections extend to a morphism from S to T. This morphism is onto since T as well is generated by its rank k elements. It is also k-injective by construction, and hence it is injective by Lemma 2.12 since S is k-determined. □

2.5 Preclones, magmoids and strict monoidal categories

The point of this short subsection is to verify the close connection between the category of preclones and the category of *magmoids*, cf. [2], which are in turn a special case of *strict monoidal categories*, cf. [28]. We recall that a magmoid is a category M whose objects are the nonnegative integers equipped with an associative bifunctor $\oplus$ such that $0 \oplus x = x = x \oplus 0$. A morphism of magmoids is a functor that preserves objects and $\oplus$. We say that a magmoid M is *determined by its scalar morphisms* if each morphism $f\colon n \to m$ can be written in a unique way as a $\oplus$-sum $f_1 \oplus \cdots \oplus f_n$, where each f_i is a morphism with source 1. Moreover, there is a morphism 0 to n if and only if $n = 0$ (in which case there is a unique morphism).

Proposition 2.16 *The category of preclones is equivalent to the full subcategory of magmoids spanned by those magmoids which are determined by their scalar morphisms.*

Proof. With each preclone S, we associate a category whose objects are the nonnegative integers and whose morphisms $n \to m$ are the elements of $S_{n,m}$,

that is, the n-tuples of elements of S of total rank m. Composition is defined in the following way: let $f = f_1 \oplus \cdots \oplus f_n \in S_{n,m}$ and $g = g_1 \oplus \cdots \oplus g_m \in S_{m,p}$, and suppose that $f_i \in S_{m_i}$, $i \in [n]$ (so that $m = \sum_{i \in [n]} m_i$). For each i, let $\overline{g}_i = g_{m_1+\cdots+m_{i-1}+1} \oplus \cdots \oplus g_{m_1+\cdots+m_i}$. Then we let

$$f \cdot g = f_1 \cdot \overline{g}_1 \oplus \cdots \oplus f_m \cdot \overline{g}_m.$$

The identity morphism at object n is the n-tuple $\mathbf{n} = \mathbf{1} \oplus \cdots \oplus \mathbf{1}$. Note that when $n = 0$, this is the unique morphism $0 \to 0$, and there are no morphisms from 0 to n if $n \neq 0$.

One may then regard $\oplus$ as a bifunctor $S \times S \to S$ that maps a pair (f, g) with $f = f_1 \oplus \cdots \oplus f_n \in S_{n,p}$ and $g = g_1 \oplus \cdots \oplus g_m \in S_{m,q}$ to the morphism $f_1 \oplus \cdots \oplus f_n \oplus g_1 \oplus \cdots \oplus g_m$ from $n + m$ to $p + q$. Then S, equipped with the bifunctor $\oplus$, is a magmoid. Moreover, S, as a magmoid, is determined by its scalar morphisms.

It is clear that each preclone morphism determines a functor between the corresponding magmoids which is the identity function on objects and preserves $\oplus$ and is thus a morphism of magmoids.

Conversely, if M is a magmoid determined by its scalar morphism, then its morphisms with sourse 1 constitute a preclone S, moreover, M is isomorphic to the magmoid determined by S. $\square$

3 Recognizable tree languages

As discussed in the introduction, the theory of (regular) tree languages is well developped [23,24,8] (see also Section 2.3.1 above). Here we slightly extend the notion of tree languages, to mean any subset of some ΣM_k, $k \geq 0$. In the classical setting, tree languages are subsets of ΣM_0.

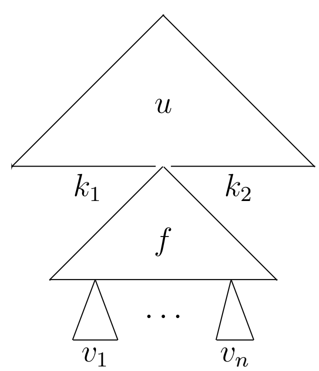
The preclone structure on ΣM, described in Section 2, leads in a standard fashion to a definition of recognizable tree languages [17,30,9,12,40]. This is discussed in some detail in Section 3.1. As we will see in Section 3.2, recognizability extends the classical notion of regularity for tree languages, and it gives us richer algebraic tools to discuss these languages. Further examples are given in Section 3.3.

Suppose that $\alpha : \Sigma M \to S$ is a preclone morphism, or a morphism $(\Sigma M, \Sigma) \to (S, A)$. We say that a subset L of ΣM_k is *recognized by* α if $L = \alpha^{-1}\alpha(L)$, or equivalently, if $L = \alpha^{-1}(F)$ for some $F \subseteq S_k$. Moreover, we say that L is recognized by S, or by (S, A), if L is recognized by some morphism $\Sigma M \to S$ or $(\Sigma M, \Sigma) \to (S, A)$. Finally, we say that a subset L of ΣM_k is *recognizable* if it is recognized by a finitary preclone, or *pg*-pair. As usual, the notion of recognizability can be expressed equivalently by stating that L is *saturated* by some *locally finite* congruence on ΣM, that is, L is a union of classes of a congruence which has finite index on each sort [10,11,40].

With every subset $L \subseteq \Sigma M_k$, recognizable or not, we associate a congruence on ΣM, called the *syntactic congruence* of L. This relation is defined as follows. First, an *n-ary context in* ΣM_k is a tuple (u, k_1, v, k_2) where

- k_1, k_2 are nonnegative integers,
- $u \in \Sigma M_{k_1+1+k_2}$, and
- $v = v_1 \oplus \cdots \oplus v_n \in \Sigma M_{n,\ell}$, with $k = k_1 + \ell + k_2$.

(u, k_1, v, k_2) is an *L-context* of an element $f \in \Sigma M_n$ if $u \cdot (\mathbf{k}_1 \oplus f \cdot v \oplus \mathbf{k}_2) \in L$. Recall that $\mathbf{k}$ denotes the $\oplus$-sum of k terms equal to $\mathbf{1}$. Below, when k_1 and k_2 are clear from the context (or do not play any role), we will write just (u, v) to denote the context (u, k_1, v, k_2).



For each $f, g \in \Sigma M_n$, we let $f \sim_L g$ if and only if f and g have the same L-contexts.

Proposition 3.1 *The relation $\sim_L$, associated with a subset L of ΣM_k, is a preclone congruence which saturates L.*

Proof. Suppose that $f, f' \in \Sigma M_n$ and $g, g' \in \Sigma M_{n,m}$ with $f \sim_L f'$, $g = g_1 \oplus \cdots \oplus g_n$, $g' = g'_1 \oplus \cdots \oplus g'_n$ and $g_i \sim_L g'_i$ for each $1 \le i \le n$. We prove that $f \cdot g \sim_L f' \cdot g'$. Let m_i be the rank of g_i and g'_i, so that $m = \sum_{i\in[n]} m_i$ and consider any m-ary context (u, k_1, v, k_2) in ΣM_k. Then $v \in \Sigma M_{m,\ell}$ with

$\ell = k - (k_1 + k_2)$. Thus, v is an m-tuple, and we let w_1 be the $\oplus$-sum of the first m_1-terms of v, w_2 be the $\oplus$-sum of the following m_2-terms of v, etc, until finally w_n is the $\oplus$-sum of the last m_n terms of v. In particular, we may write $v = w_1 \oplus \cdots \oplus w_n$.

Since $f \sim_L f'$, we have

$$u \cdot (\mathbf{k}_1 \oplus f \cdot g \cdot v \oplus \mathbf{k}_2) \in L \Longleftrightarrow u \cdot (\mathbf{k}_1 \oplus f' \cdot g \cdot v \oplus \mathbf{k}_2) \in L.$$

It suffices to consider the n-ary context $(u, \mathbf{k}_1, g \cdot v, \mathbf{k}_2)$, where $g \cdot v = (g_1 \oplus \cdots \oplus g_n) \cdot v$ stands for $g_1 \cdot w_1 \oplus \cdots \oplus g_n \cdot w_n$.

Moreover, since $g_i \sim_L g'_i$, we have

$$\begin{aligned} & u \cdot (\mathbf{k}_1 \oplus f \cdot (g'_1 \cdot w_1 \oplus \cdots \oplus g'_{i-1} \cdot w_{i-1} \oplus g_i \cdot w_i \oplus g_{i+1} \cdot w_{i+1} \cdots \oplus g_n \cdot w_n) \oplus \mathbf{k}_2) \in L \\ \Longleftrightarrow\, & u \cdot (\mathbf{k}_1 \oplus f \cdot (g'_1 \cdot w_1 \oplus \cdots \oplus g'_i \cdot w_i \oplus g_{i+1} \cdot w_{i+1} \oplus \cdots \oplus g_n \cdot w_n) \oplus \mathbf{k}_2) \in L, \end{aligned}$$

for each $1 \leq i \leq n$. To justify this statement, it suffices to consider the following m_i-ary context (for g_i and g'_i),

$$\Big(u \cdot (g'_1 \cdot w_1 \oplus \cdots g'_{i-1} \cdot w_{i-1} \oplus \mathbf{1} \oplus g_{i+1} \cdot w_{i+1} \oplus \cdots \oplus g_n \cdot w_n),\ \mathbf{k}_1 + \mathbf{l}_{i,1},\ w_i,\ \mathbf{l}_{i,2} + \mathbf{k}_2\Big),$$

where $\ell_{i,1}$ is the sum of the ranks of $w_1, \ldots, w_{i-1}$, and $\ell_{i,2}$ is the sum of the ranks of $w_{i+1}, \ldots, w_n$.

We now have

$$\begin{aligned} u \cdot (\mathbf{k}_1 \oplus f' \cdot g \cdot v \oplus \mathbf{k}_2) \in L & \Longleftrightarrow u \cdot (\mathbf{k}_1 \oplus f \cdot (g_1 \cdot w_1 \oplus \cdots \oplus g_n \cdot w_n) \oplus \mathbf{k}_2) \in L \\ & \Longleftrightarrow u \cdot (\mathbf{k}_1 \oplus f \cdot (g'_1 \cdot w_1 \oplus \cdots \oplus g_n \cdot w_n) \oplus \mathbf{k}_2) \in L \\ & \quad \vdots \\ & \Longleftrightarrow u \cdot (\mathbf{k}_1 \oplus f \cdot (g'_1 \cdot w_1 \oplus \cdots \oplus g'_n \cdot w_n) \oplus \mathbf{k}_2) \in L \\ & \Longleftrightarrow u \cdot (\mathbf{k}_1 \oplus f \cdot g' \cdot v \oplus \mathbf{k}_2) \in L. \end{aligned}$$

This completes the proof that $\sim_L$ is a congruence. Next we observe that an element $f \in \Sigma M_k$ is in L if and only if the k-ary context $(\mathbf{1}, \mathbf{k})$ is an L-context of f: it follows immediately that $\sim_L$ saturates L. □

We denote by (M_L, Σ_L) the quotient *pg*-pair $(\Sigma M/{\sim_L}, \Sigma/{\sim_L})$, called the *syntactic pg-pair* of L. M_L is the *syntactic preclone* of L and the projection morphism $\eta_L \colon \Sigma M \to M_L$, or $\eta_L \colon (\Sigma M, \Sigma) \to (M_L, \Sigma_L)$, is the *syntactic morphism* of L. We note the following, expected result.

Proposition 3.2 *The syntactic congruence of a subset L of ΣM_k is the coarsest preclone congruence which saturates L. A preclone morphism $\alpha \colon \Sigma M \to S$*

(resp. a morphism of pg-pairs $\alpha\colon (\Sigma M, \Sigma) \to (S, A)$) recognizes L if and only if α can be factored through the syntactic morphism η_L. In particular, L is recognizable if and only if $\sim_L$ is locally finite, if and only if M_L is finitary.

Proof. Let $\approx$ be a congruence saturating L and assume that $f, g \in \Sigma M_n$ are $\approx$-equivalent. Let (u, k_1, v, k_2) be an n-ary context: then $u \cdot (\mathbf{k}_1 \oplus f \cdot v \oplus \mathbf{k}_2) \approx u \cdot (\mathbf{k}_1 \oplus g \cdot v \oplus \mathbf{k}_2)$, and since $\approx$ saturates L, $u \cdot (\mathbf{k}_1 \oplus f \cdot v \oplus \mathbf{k}_2) \in L$ iff $u \cdot (\mathbf{k}_1 \oplus g \cdot v \oplus \mathbf{k}_2) \in L$. Since this holds for all n-ary contexts in ΣM_k, it follows that $f \sim_L g$. $\square$

We also note that syntactic preclones are finitely determined.

Proposition 3.3 *The syntactic preclone of a subset L of ΣM_k is k-determined.*

Proof. We show that if $f, g \in \Sigma M_n$ and $f \cdot h \sim_L g \cdot h$ for all $h \in \Sigma M_{n,\ell}$ with $\ell \leq k$, then $f \sim_L g$, that is, f and g have the same L-contexts.

Let (u, k_1, v, k_2) be an L-context of f. Note in particular that $v \in \Sigma M_{n,p}$ with $k = k_1 + p + k_2$. It follows that $f \cdot v \in \Sigma M_p$, and that $f \cdot v \sim_L g \cdot v$. Moreover $(u, k_1, \mathbf{p}, k_2)$ is an L-context of $f \cdot v$. But in that case, $(u, k_1, \mathbf{p}, k_2)$ is also an L-context of $g \cdot v$, and hence (u, k_1, v, k_2) is an L-context of g, which concludes the proof. $\square$

3.2 The usual notion of regular tree languages

We now turn to tree languages in the usual sense, that is, subsets of ΣM_0. For these sets, there exists a well-known notion of (bottom-up) tree automaton, whose expressive power is equivalent to monadic second-order definability, to certain rational expressions, and to recognizability by a finite Σ-algebra [23,24] (see Section 2.3.1). The tree languages captured by these mechanisms are said to be *regular*. It is an essential remark (Theorem 3.4 below) that the regular tree languages are exactly the subsets of ΣM_0 that are recognized by a finitary preclone.

Recall that the minimal tree automaton of a regular tree language is the least deterministic tree automaton accepting it, and the Σ-algebra associated with this automaton is called the *syntactic Σ-algebra* of the language. It is characterized by the fact that the natural morphism from the initial Σ-algebra to the syntactic Σ-algebra of L, factors through every morphism of Σ-algebra which recognizes L (see Section 2.3.1 and [23,24,1]).

Theorem 3.4 *A tree language $L \subseteq \Sigma M_0$ is recognizable if and only if it is regular. Moreover, the syntactic preclone (resp. pg-pair) of L is the preclone*

(resp. pg-pair) associated with its syntactic Σ-algebra.

Proof. Let Q be the syntactic Σ-algebra of L, let (S, A) be its syntactic *pg*-pair, and let $\eta \colon (\Sigma M, \Sigma) \to (S, A)$ be its syntactic morphism. As discussed in Section 2.3.1, the *pg*-pair associated with Q, $\mathsf{pg}(Q)$, recognizes L, and hence the syntactic morphism of L factors through an onto morphism of *pg*-pairs $\mathsf{pg}(Q) \to (S, A)$. In particular, if L is regular, then Q is finite, so $\mathsf{pg}(Q)$ is finitary, and so is (S, A): thus L is recognizable.

Conversely, assume that L is recognizable. Since (S, A) is finitary and 0-determined (Proposition 3.3), so (S, A) is isomorphic to a sub-*pg*-pair of $T(S_0)$ by Proposition 2.13. Using again the discussion in Section 2.3.1, S_0 has a natural structure of Σ-algebra (via the morphism η), such that $(S, A) = \mathsf{pg}(S_0)$ and such that S_0 recognizes L as a Σ-algebra. In particular, L is recognized by a finite Σ-algebra, and hence L is regular.

Moreover, the recognizing morphism $\Sigma M_0 \to S_0$ is the restriction to ΣM_0 of η, the syntactic morphism of L. Therefore there exists an onto morphism of Σ-algebras $S_0 \to Q$, which in turn induces a morphism of preclones from $(S, A) = \mathsf{pg}(S_0)$ onto $\mathsf{pg}(Q)$. Since Q and S_0 are finite, it follows that the morphisms between them described above are isomorphisms, and this implies that $\mathsf{pg}(Q)$ is isomorphic to (S, A). $\square$

While not difficult, Theorem 3.4 is important because it shows that we are not introducing a new class of recognizable tree languages. We are simply associating with each regular tree language a finitary algebraic structure which is richer than its syntactic Σ-algebra (a.k.a. minimal deterministic tree automaton). This theorem also implies that the syntactic *pg*-pair of a recognizable tree language has an effectively computable finite presentation.

Remark 3.5 If $L \subseteq \Sigma M_0$, the definition of the syntactic congruence of L involves the consideration of n-ary contexts in ΣM_0. Such contexts are necessarily of the form $(u, \mathbf{0}, v, \mathbf{0})$, where $u \in \Sigma M_1$ and $v \in \Sigma M_{n,0}$, which somewhat simplifies matters. $\square$

3.3 More examples of recognizable tree languages

The examples in this section are directly related with the preclones discussed in Section 2.3.2. Let Δ be a ranked *Boolean* alphabet, that is, a ranked alphabet such that each Δ_n is either empty or equal to $\{0_n, 1_n\}$, and Δ_0 and at least one Δ_n $(n \geq 2)$ are nonempty. Let $k \geq 0$ be an integer.

Verifying the occurrence of a letter Let $K_k(\exists)$ be the set of all trees in ΔM_k containing at least one vertex labeled 1_n (for some n). Then $K_k(\exists)$ is recognizable, by a morphism into the preclone $T_\exists$ (see Example 2.5).

Let $\alpha\colon \Delta M \to T_\exists$ be the morphism of preclones given by $\alpha(0_n) = \mathsf{or}_n$ ($\alpha(0_0) = \mathsf{false}_0$) and $\alpha(1_n) = \mathsf{true}_n$ whenever $\Delta_n \neq \emptyset$. It is not difficult to verify that $\alpha^{-1}(\mathsf{true}_k) = K_k(\exists)$. Moreover, $\alpha(\Delta)$ contains a generating set of $T_\exists$, so α is onto, and the syntactic morphism of $K_k(\exists)$ factors through α. But $T_\exists$ has at most 2 elements of each rank, so any proper quotient M of $T_\exists$ has exactly one element of rank n for some integer n. One can then show that M cannot recognize $K_k(\exists)$. Thus the syntactic pg-pair of $K_k(\exists)$ is $(T_\exists, \alpha(\Delta))$.

If Σ is any ranked alphabet such that Σ_0 and at least one Σ_n ($n > 1$) is nonempty, if Σ' is a proper nonempty subset of Σ, and $K_k(\Sigma')$ is the set of all trees in $\Sigma' M_k$ containing at least a node labeled in Σ', then $K_k(\Sigma')$ too has syntactic preclone $T_\exists$. The verification of this fact can be done using a morphism from ΣM to ΔM, mapping each letter σ of rank n to 1_n if it is in Σ', to 0_n otherwise.

Counting the occurrences of a letter Let p, r be integers such that $0 \leq r < p$ and let $K_k(\exists_p^r)$ consist of the trees in ΔM_k such that the number of vertices labeled 1_n (for some n) is congruent to r modulo p. Then $K_k(\exists_p^r)$ is recognizable, by a morphism into the preclone T_p (see Example 2.6).

Let indeed $\alpha\colon \Delta M \to T_p$ be the morphism given by $\alpha(0_n) = f_{n,0}$ and $\alpha(1_n) = f_{n,1}$ whenever $\Delta_n \neq \emptyset$. Then one verifies that $\alpha^{-1}(f_{k,r}) = K_k(\exists_p^r)$. Moreover, $\alpha(\Delta)$ contains a generating set of T_p, so α is onto, and the syntactic morphism of $K_k(\exists_p^r)$ factors through α. An elementary verification then establishes that no proper quotient of T_p can recognize $K_k(\exists_p^r)$, and hence the syntactic pg-pair of $K_k(\exists_p^r)$ is $(T_p, \alpha(\Delta))$.

As above, this can be extended to recognizing the set of all trees in ΣM_k where the number of nodes labeled in some proper nonempty subset Σ' of Σ is congruent to r modulo p.

Using the same idea, one can also handle tree languages defined by counting the number of occurrences of certain letters modulo p threshold q. It suffices to consider, in analogy with the mod p case, the languages of the form $K_k(\exists_{p,q}^r)$, and the preclone $T_{p,q}$, a sub-preclone of $T(\mathbb{B}_{p+q})$, whose rank n elements are the mappings $f_r\colon (r_1, \ldots, r_n) \mapsto r_1 + \cdots + r_n + r$, where the sum is taken modulo p threshold q. Note that this notion generalizes both above examples, since $T_p = T_{p,0}$ and that $T_\exists = T_{1,1}$.

Identification of a path Let $K_k(\mathsf{path})$ be the set of all the trees in ΔM_k such that all the vertices along at least one maximal path from the root to a leaf are labeled 1_n (for the appropriate values of n). Then $K_k(\mathsf{path})$ is recognized by the preclone T_{path} (see Example 2.7).

Let indeed $\alpha\colon \Delta M \to T_{\mathsf{path}}$ be the morphism given by $\alpha(0_n) = \mathsf{false}_n$, $\alpha(1_0) = \mathsf{true}_0$ and $\alpha(1_n) = \mathsf{or}_n$ ($n \neq 0$). One can then verify that $\alpha^{-1}(\mathsf{true}_k) = K_k(\mathsf{path})$.

Identification of the next modality Let $K_k(\mathsf{next})$ consist of all the trees in ΔM_k such that each maximal path has length at least two and the children of the root are labeled 1_n (for the appropriate n). We show that $K_k(\mathsf{next})$ is recognizable.

Recall that $\mathbb{B} = \{\mathsf{true}, \mathsf{false}\}$, and let $\alpha\colon \Delta M \to T(\mathbb{B} \times \mathbb{B})$ be the morphism given as follows:

- $\alpha(0_0)$ is the nullary constant $(\mathsf{false}, \mathsf{false})_0$,
- $\alpha(1_0)$ is the nullary constant $(\mathsf{false}, \mathsf{true})_0$,
- if $n > 0$, then $\alpha(0_n)$ is the n-ary map $((x_1, y_1), \ldots, (x_n, y_n)) \mapsto (\wedge_i y_i, \mathsf{false})$
- if $n > 0$, then $\alpha(1_n)$ is the n-ary map $((x_1, y_1), \ldots, (x_n, y_n)) \mapsto (\wedge_i y_i, \mathsf{true})$.

One can verify by structural induction that for each element $x \in \Delta M_k$, the second component of $\alpha(x)$ is true if and only if the root of x is labeled 1_n for some n, and the first component of $\alpha(x)$ is true if and only if every child of the root of x is labeled 1_n for some n, that is, if and only if $x \in K_k(\mathsf{next})$. Thus $K_k(\mathsf{next})$ is recognized by the morphism α.

4 Pseudovarieties of preclones

In the usual setting of one-sorted algebras, a pseudovariety is a class of finite algebras closed under taking finite direct products, sub-algebras and quotients. Because we are dealing with preclones, which are infinitely sorted, we need to consider finitary algebras instead of finite ones, and to adopt more constraining closure properties in the definition. (We discuss in Remark 4.18 an alternative approach, which consists in introducing stricter finiteness conditions on the preclones themselves, namely in considering only finitely generated, finitely determined, finitary preclones.)

We say that a class of finitary preclones is a *pseudovariety* if it is closed under finite direct product, sub-preclones, quotients, finitary unions of ω-chains and finitary inverse limits of ω-diagrams. Here, we say that a union $T = \bigcup_n T^{(n)}$

of an ω-chain of preclones $T^{(n)}$, $n \geq 0$ is finitary exactly when T is finitary. Finitary inverse limits $\lim_n T^{(n)}$ of ω-diagrams $\varphi_n: T^{(n+1)} \to T^{(n)}$, $n \geq 0$ are defined in the same way.

Remark 4.1 To be perfectly rigorous, we actually require pseudovarieties to be closed under taking preclones *isomorphic to* a finitary ω-union or to a finitary inverse limit of an ω-diagram of their elements. $\square$

Remark 4.2 Recall that the inverse limit T of the ω-diagram $(\varphi_n)_{n\geq 0}$, written $T = \lim_n T^{(n)}$ if the $\varphi_n: T^{(n+1)} \to T^{(n)}$ are clear, is the sub-preclone of the direct product $\prod_n T^{(n)}$ whose set of elements of rank m consists of those sequences $(x_n)_{n\geq 0}$ with $x_n \in T^{(n)}_m$ such that $\varphi_n(x_{n+1}) = x_n$, for all $n \geq 0$. We call the coordinate projections $\pi_p : \lim_n T^{(n)} \to T^{(p)}$ the *induced projection morphisms*.

$$
\begin{tikzcd}
& & T \arrow[d, "\pi_{n+1}"'] \arrow[dr, "\pi_n"] & & & \\
& \cdots & & & \cdots & \\
\cdots \arrow[r] & & T^{(n+1)} \arrow[r, "\varphi_n"'] & T^{(n)} \arrow[r] & \cdots \arrow[r] & T^{(0)}
\end{tikzcd}
$$

The inverse limit has the following universal property. Whenever S is a preclone and the morphisms $\psi_n: S \to T^{(n)}$ satisfy $\psi_n = \varphi_n \circ \psi_{n+1}$ for each $n \geq 0$, then there is a unique morphism $\psi: S \to \lim_n T^{(n)}$ with $\pi_n \circ \psi = \psi_n$, for all n. This morphism ψ maps an element $s \in S$ to the sequence $(\psi_n(s))_{n\geq 0}$. $\square$

Example 4.3 Here we show that the inverse limit of an ω-diagram of 1-generated finitary preclones needs not be finitary. Let $\Sigma = \{\sigma\}$, where σ has rank 1 and consider the free preclone ΣM. Note that ΣM has only elements of rank 1, and that ΣM_1 can be identified with the monoid σ^*. For each $n \geq 0$, let $\approx_n$ be the congruence defined by letting $\sigma^k \approx_n \sigma^\ell$ if and only if $k = \ell$, or $k, \ell \geq n$. Let $T^{(n)} = \Sigma M/\!\approx_n$. Then $T^{(n)}$ is again σ-generated, and it can be identified with the monoid $\{0, 1, \ldots, n\}$ under addition threshold n. In particular, $T^{(n)}$ is a finitary preclone. Since $\approx_{n+1}$-equivalent elements of ΣM are also $\approx_n$-equivalent, there is a natural morphism of preclones from $T^{(n+1)}$ to $T^{(n)}$, mapping σ to itself, and the inverse limit of the resulting ω-diagram is ΣM itself, which is not finitary. $\square$

Pseudovarieties of preclones can be characterized using the notion of *division*: we say that a preclone S *divides* a preclone T, written $S < T$, if S is a quotient of a sub-preclone of T. It is immediately verified that a nonempty class of finitary preclones is a pseudovariety if and only if it is closed with respect to division, binary direct product, finitary unions of ω-chains and finitary inverse limits of ω-diagrams.

Example 4.4 It is immediate that the intersection of a collection of pseudovarieties of preclones is a pseudovariety. It follows that if $\mathbf{K}$ is a class of finitary preclones, then the pseudovariety generated by $\mathbf{K}$ is well defined, as the least pseudovariety containing $\mathbf{K}$. In particular, the elements of this pseudovariety, written $\langle \mathbf{K} \rangle$, can be described in terms of the elements of $\mathbf{K}$, taking subpreclones, quotients, direct products, finitary unions of ω-chains and inverse limits of ω-diagrams. See Section 4.2 below.

We discuss other examples in Section 5.2. $\square$

We first explore the relation between pseudovarieties and their finitely determined elements, then we discuss pseudovarieties generated by a class of preclones, and finally, we explore some additional closure properties of pseudovarieties.

4.1 Pseudovarieties and their finitely determined elements

Proposition 4.5 *Let S be a preclone.*

(1) S is isomorphic to the inverse limit $\lim_n S^{(n)}$ of an ω-diagram, where each $S^{(n)}$ is an n-determined quotient of S.
(2) If S is finitary, then S is isomorphic to the union of an ω-chain $\bigcup_{n\geq 0} T^{(n)}$, where each $T^{(n)}$ is the inverse limit of an ω-diagram of finitely generated, finitely determined divisors of S.

Proof. Let $S^{(n)} = S/{\sim_n}$ (where $\sim_n$ is defined in Section 2.4) and let $\pi_n\colon S \to S^{(n)}$ be the corresponding projection. Since $\sim_{n+1}$-related elements of S are also $\sim_n$-related, there exists a morphism of preclones $\varphi_n\colon S^{(n+1)} \to S^{(n)}$ such that $\pi_n = \varphi_n \circ \pi_{n+1}$. Thus the π_n determine a morphism $\pi\colon S \to \lim_n S^{(n)}$, such that $\pi(s) = (\pi_n(s))_n$ for each $s \in S$ (Remark 4.2).

Moreover, since $\sim_n$ is the identity relation on the elements of S of rank at most n, we find that for each $k \leq n$, π_n establishes a bijection between the elements of rank k of S and those of $S^{(n)}$. In particular, π is injective since each element of S has rank k for some finite integer k. Furthermore, for each $k \leq n$, φ_n establishes a bijection between the elements of rank k, and it follows that each element of rank k of $\lim_n S^{(n)}$ is the π-image of its k-th component. That is, π is onto. Finally, Lemma 2.11 shows that each $S^{(n)}$ is n-determined. This concludes the proof of the first statement.

We now assume that S is finitary, and we let $T^{(m)}$ be the sub-preclone generated by the elements of S of rank at most m. Then $T^{(m)}$ is finitely generated, and the first statement shows that $T^{(m)}$ is the inverse limit of an ω-diagram

of finitely generated, finitely determined quotients of $T^{(m)}$, which are in particular divisors of S. □

The following corollary follows immediately.

Corollary 4.6 *Every pseudovariety of preclones is uniquely determined by its finitely generated, finitely determined elements.*

We can go a little further, and show that a pseudovariety is determined by the syntactic preclones it contains.

Proposition 4.7 *Let S be a finitely generated, k-determined, finitary preclone, let A be a finite ranked set and let $\varphi\colon AM \to S$ be an onto morphism. Then S divides the direct product of the syntactic preclones of the languages $\varphi^{-1}(s)$, where s runs over the (finitely many) elements of S of rank at most k.*

Proof. It suffices to show that if $x, y \in AM_n$ for some $n \geq 0$ and $x \sim_{\varphi^{-1}(s)} y$ for each $s \in S_\ell$, $\ell \leq k$, then $\varphi(x) = \varphi(y)$.

First, suppose that x and y have rank $n \leq k$, and let $s = \varphi(x)$. Then $(\mathbf{1}, 0, \mathbf{n}, 0)$ is a $\varphi^{-1}(s)$-context of x, so it is also a $\varphi^{-1}(s)$-context of y, and we have $\varphi(y) = s = \varphi(x)$. Now, if x and y have rank $n > k$, let $v \in S_{n,p}$ for some $p \leq k$. Since φ is onto, there exists an element $z \in AM_{n,p}$ such that $\varphi(z) = v$. For each $s \in S_\ell$, $\ell \leq k$, we have $x \sim_{\varphi^{-1}(s)} y$, and hence also $x \cdot z \sim_{\varphi^{-1}(s)} y \cdot z$. The previous discussion shows therefore that $\varphi(x \cdot z) = \varphi(y \cdot z)$, that is, $\varphi(x) \cdot v = \varphi(y) \cdot v$. Since S is k-determined, it follows that $\varphi(x) = \varphi(y)$. □

Corollary 4.8 *Every pseudovariety of preclones is uniquely determined by the syntactic preclones it contains.*

Proof. This follows directly from Corollary 4.6 and Proposition 4.7. □

4.2 The pseudovariety generated by a class of preclones

Let $\mathbf{I}, \mathbf{H}, \mathbf{S}, \mathbf{P}, \mathbf{L}, \mathbf{U}$ denote respectively the operators of taking all isomorphic images, homomorphic images, subpreclones, finite direct products, finitary inverse limits of an ω-diagram, and finitary ω-unions over a class of finitary preclones. The following fact is a special case of a well-known result in universal algebra.

Lemma 4.9 *If $\mathbf{K}$ is a class of finitary preclones, then $\mathbf{HSP}(\mathbf{K})$ is the least class of finitary preclones containing $\mathbf{K}$, closed under homomorphic images,*

subpreclones and finite direct products.

Next, we observe the following elementary facts.

Lemma 4.10 *For all classes* $\mathbf{K}$ *of finitary preclones, we have*

$$(1)\ \mathbf{PL}(\mathbf{K}) \subseteq \mathbf{LP}(\mathbf{K}), \qquad (2)\ \mathbf{PU}(\mathbf{K}) \subseteq \mathbf{UP}(\mathbf{K}),$$
$$(3)\ \mathbf{SL}(\mathbf{K}) \subseteq \mathbf{LS}(\mathbf{K}), \qquad (4)\ \mathbf{SU}(\mathbf{K}) \subseteq \mathbf{US}(\mathbf{K}).$$

Proof. To prove the first inclusion, suppose that S is the direct product of the finitary preclones $S^{(i)}$, $i \in [n]$, where each $S^{(i)}$ is a limit of an ω-diagram of preclones $S^{(i,k)}$ in $\mathbf{K}$ determined by a family of morphisms $\varphi_{i,k}\colon S^{(i,k+1)} \to S^{(i,k)}$, $k \geq 0$. For each k, let $T^{(k)}$ be the direct product $\prod_{i\in[n]} S^{(i,k)}$, and let $\varphi_k = \prod_{i\in[n]} \varphi_{i,k}\colon T^{(k+1)} \to T^{(k)}$. It is a routine matter to verify that S is isomorphic to the limit of the ω-diagram determined by the family of morphisms $\varphi_k\colon T^{(k+1)} \to T^{(k)}$, $k \geq 0$. Thus, $S \in \mathbf{LP}(\mathbf{K})$.

Now, for each $i \in [n]$, let $(S^{(i,k)})_{k\geq 0}$ be an ω-chain of finitary preclones in $\mathbf{K}$. Let us assume that each $S^{(i)} = \bigcup_{k\geq 0} S^{(i,k)}$ is finitary, and let $S = \prod_{i\in[n]} S^{(i)}$. If $s = (s_1, \ldots, s_n) \in S$, then each s_i belongs to $S^{(i,k_i)}$, for some k_i. Thus $s \in \prod_{i\in[n]} S^{(i,k)}$, where $k = \max k_i$, and we have shown that $S \in \bigcup_{k\geq 0} \prod_{i\in[n]} S^{(i,k)}$, so that $S \in \mathbf{UP}(\mathbf{K})$.

To prove the third inclusion, let T be a sub-preclone of $\lim_n S^{(n)}$, the finitary inverse limit of an ω-diagram $\varphi_n\colon S^{(n+1)} \to S^{(n)}$ of elements of $\mathbf{K}$. Let $\pi_n\colon T \to S^{(n)}$ be the natural projections (restricted to T), and let $T^{(n)} = \pi_n(T)$. Then $T^{(n)}$ is a subpreclone of $S^{(n)}$ for each n. Moreover, the restrictions of the φ_n to $T^{(n+1)}$ define an ω-diagram of subpreclones of elements of $\mathbf{K}$, and it is an elementary verification that $T = \lim_n T^{(n)}$. Since T is finitary, we have proved that $T \in \mathbf{LS}(\mathbf{K})$.

As for the last inclusion, let T be a subpreclone of a finitary union $\bigcup_{k\geq 0} S^{(k)}$ with $S^{(k)} \in \mathbf{K}$, for all $k \geq 0$. Let $T^{(k)} = S^{(k)} \cap T$ for each $k \geq 0$. Then each $T^{(k)}$ is a subpreclone of $S^{(k)}$ and $T = \bigcup_{k\geq 0} T^{(k)}$. It follows that $T \in \mathbf{US}(\mathbf{K})$. □

Our proof of the third inclusion actually yields the following result.

Corollary 4.11 *If a finitary preclone* S *embeds in an inverse limit* $\lim_n S^{(n)}$*, then* S *is isomorphic to a (finitary) inverse limit* $\lim_n T^{(n)}$*, where each* $T^{(n)}$ *is a finitary sub-preclone of* $S^{(n)}$*.*

We can be more precise than Lemma 4.10 for what concerns finitely generated, finitely determined preclones.

Lemma 4.12 *Let* T *be a preclone which embeds in the union of an* ω*-chain*

$(S^{(n)})_n$. *If T is finitely generated, then T embeds in $S^{(n)}$ for all large enough n.*

Proof. Since T is finitely generated, its set of generators is entirely contained in some $S^{(k)}$, and hence T embeds in each $S^{(n)}$, $n \geq k$. $\square$

Lemma 4.13 *Let T be a quotient of the union of an ω-chain $(S^{(n)})_n$. If T is finitely generated, then T is a quotient of $S^{(n)}$ for all large enough n.*

Proof. Let φ be a surjective morphism from $S = \bigcup_n S^{(n)}$ onto T. Since T is finitely generated, there exists an integer k such that $\varphi(S^{(k)})$ contains all the generators of T, and this implies that the restriction of φ to $S^{(k)}$ (and to each $S^{(n)}$, $n \geq k$) is onto. $\square$

Lemma 4.14 *Let T be a preclone which embeds in the inverse limit* $\lim_n S^{(n)}$ *of an ω-diagram, and for each n, let $\pi_n\colon T \to S^{(n)}$ be the natural projection (restricted to T). If T is finitary, then for each k, π_n is k-injective for all large enough n. If in addition T is finitely determined, then T embeds in S_n for all large enough n.*

Proof. Since T is finitary, T_k is finite for each integer k, and hence there exists an integer n_k such that π_n is injective on T_k for each $n \geq n_k$. In particular, for each integer k, π_n is k-injective for all large enough n. The last part of the statement follows from Lemma 2.12. $\square$

Lemma 4.15 *Let T be a quotient of the finitary inverse limit* $\lim_n S^{(n)}$ *of an ω-diagram. If T is finitely determined, then T is a quotient of a sub-preclone of one of the $S^{(n)}$.*

Proof. Let $S = \lim_n S^{(n)}$ and let $\pi_n\colon S \to S^{(n)}$ be the corresponding projection. Let also $\varphi\colon S \to T$ be an onto morphism, and let $k \geq 0$ be an integer such that T is k-determined. By Lemma 4.14, π_n is k-injective for some integer n.

Consider the preclone $\pi_n(S) \subseteq S^{(n)}$. Then we claim that the assignment $\pi_n(s) \mapsto \varphi(s)$ defines a surjective morphism $\pi_n(S) \to T$. The only nontrivial point is to verify that this assignment is well defined. Let $s, s' \in S_p$ and suppose that $\pi_n(s) = \pi_n(s')$. We want to show that $\varphi(s) = \varphi(s')$, and for that purpose, we show that $\varphi(s) \cdot v = \varphi(s') \cdot v$ for each $v \in T_{p,\ell}$, $\ell \leq k$ (since T is k-determined). Since φ is onto, there exists $w \in S_{p,\ell}$ such that $v = \varphi(w)$. In particular, $\varphi(s) \cdot v = \varphi(s \cdot w)$ and similarly, $\varphi(s') \cdot v = \varphi(s' \cdot w)$. Moreover, we have $\pi_n(s \cdot w) = \pi_n(s' \cdot w)$. Now $s \cdot w$ and $s' \cdot w$ lie in S_ℓ, and π_n is injective on S_ℓ, so $s \cdot w = s' \cdot w$. It follows that $\varphi(s) \cdot v = \varphi(s') \cdot v$, and hence $\varphi(s) = \varphi(s')$. $\square$

We are now ready to describe the finitely generated, finitely determined elements of the pseudovariety generated by a given class of finitary preclones.

Proposition 4.16 *Let* $\mathbf{K}$ *be a class of finitary preclones. A finitely generated, finitely determined, finitary preclone belongs to the pseudovariety* $\langle\mathbf{K}\rangle$ *generated by* $\mathbf{K}$ *if and only if it divides a finite direct product of preclones in* $\mathbf{K}$*, i.e., it lies in* $\mathbf{HSP}(\mathbf{K})$*.*

Proof. It is easily verified that $\langle\mathbf{K}\rangle = \bigcup_n \mathbf{V}_n$, where $\mathbf{V}_0 = \mathbf{K}$ and $\mathbf{V}_{n+1} = \mathbf{HSPUHSPL}(\mathbf{V}_n)$. We show by induction on n that if T a finitely generated, finitely determined preclone in $\mathbf{V}_n$, then $T \in \mathbf{HSP}(\mathbf{K})$.

The case $n = 0$ is trivial and we now assume that $T \in \mathbf{V}_{n+1}$. By Lemma 4.10, T lies in $\mathbf{HUSPHLSP}(\mathbf{V}_n)$. Then Lemma 4.13 shows that T is in fact in $\mathbf{HSPHLSP}(\mathbf{V}_n)$, which is equal to $\mathbf{HSPLSP}(\mathbf{V}_n)$ by Lemma 4.9, and is contained in $\mathbf{HLSP}(\mathbf{V}_n)$ by Lemma 4.10 again. Now Lemma 4.15 shows that T lies in fact in $\mathbf{HSP}(\mathbf{V}_n)$, and we conclude by induction that $T \in \mathbf{HSP}(\mathbf{K})$. $\square$

Corollary 4.17 *If* $\mathbf{K}$ *is a class of finitary preclones, then* $\langle\mathbf{K}\rangle = \mathbf{IULHSP}(\mathbf{K})$*.*

Proof. The containment $\mathbf{IULHSP}(\mathbf{K}) \subseteq \langle\mathbf{K}\rangle$ is immediate. To show the reverse inclusion, we consider a finitary preclone $T \in \langle\mathbf{K}\rangle$. Then $T = \bigcup T^{(n)}$, where $T^{(n)}$ denotes the subpreclone of T generated by the elements of rank at most n. Now each $T^{(n)}$ is finitely generated, and by Proposition 4.5, it is isomorphic to the inverse limit of the ω-diagram formed by the finitely generated, finitely determined preclones $T_n/{\sim_m}$, $m \geq 0$. By the Proposition 4.16, each of these preclones is in $\mathbf{HSP}(\mathbf{K})$, so $T \in \mathbf{IULHSP}(\mathbf{K})$. $\square$

Remark 4.18 As indicated in the first paragraph of Section 4, Proposition 4.16 hints at an alternative treatment of the notion of pseudovarieties of preclones, limited to the consideration of finitely generated, finitely determined, finitary preclones. Say that a class $\mathbf{K}$ of finitely generated, finitely determined, finitary preclones is a *relative pseudovariety* if whenever a finitely generated, finitely determined, finitary preclone S divides a finite direct product of preclones in $\mathbf{K}$, then S is in fact in $\mathbf{K}$. For each pseudovariety $\mathbf{V}$, the class $\mathbf{V}_{\mathsf{fin}}$ of all its finitary, finitely generated, finitely determined members is a relative pseudovariety, and the map $\mathbf{V} \mapsto \mathbf{V}_{\mathsf{fin}}$ is injective by Corollary 4.6. Moreover, Proposition 4.16 can be used to show that this map is onto. That is, the map $\mathbf{V} \mapsto \mathbf{V}_{\mathsf{fin}}$ is an order-preserving bijective correspondence (with respect to the inclusion order) between pseudovarieties and relative pseudovarieties of preclones. $\square$

Proposition 4.16 also leads to the following useful result. Recall that a finitely generated preclone S is effectively given if we are given a finite generating set

A as transformations of finite arity of a given finite set Q, see Section 2.3.1.

Corollary 4.19 *Let S and T be effectively given, finitely generated, finitely determined preclones. Then it is decidable whether T belongs to the pseudovariety of preclones generated by S.*

Proof. Let A (resp. B) be the given set of generators of S (resp. T) and let $\mathbf{V}$ be the pseudovariety generated by S. By Corollary 4.16, $T \in \mathbf{V}$ if and only if T divides a direct power of S, say, $T < S^m$. Since B is finite, almost all the sets B_k are empty. We claim that the exponent m can be bounded by

$$\prod_{B_k \neq \emptyset} |A_k|^{|B_k|}.$$

Indeed, there exists a sub-preclone $S' \subseteq S^m$ and an onto morphism $S' \to T$. Since B generates T, we may assume without loss of generality that this morphism defines a bijection from a set A' of generators of S' to B, and in particular, we may identify B_k with A'_k, a subset of A_k^m. Next, one verifies that if m is greater than the bound in the claim, then there exist $1 \leq i < j \leq m$ such that for all k and $x \in A'_k$, the i-th and the j-th components of x are equal — but this implies that the exponent can be decreased by 1.

Thus, it suffices to test whether or not T divides S^m, where m is given by the above formula. But as discussed above, this holds if and only if A^m contains a set A' and a rank preserving bijection from A' to B which can be extended to a morphism from the sub-preclone of S^m generated by A' to T. By Proposition 2.14, and since S and T are effectively given and T is finitely determined, this can be checked algorithmically. □

4.3 Closure properties of pseudovarieties

Here we record additional closure properties of pseudovarieties of preclones.

Lemma 4.20 *Let $\mathbf{V}$ be a pseudovariety of preclones and let T be a finitary preclone. If T embeds in the inverse limit of an ω-diagram of preclones in $\mathbf{V}$, then $T \in \mathbf{V}$.*

Proof. The lemma follows immediately from Corollary 4.11. □

Proposition 4.21 *Let $\mathbf{V}$ be a pseudovariety of preclones and let S be a finitary preclone. If for each $n \geq 0$, there exists a morphism $\varphi_n \colon S \to S^{(n)}$ such that $S^{(n)} \in \mathbf{V}$ and φ_n is injective on elements of rank exactly n, then $S \in \mathbf{V}$.*

Proof. Without loss of generality we may assume that each φ_n is surjective.

For each $n \geq 0$, consider the direct product $T^{(n)} = S^{(0)} \times \cdots \times S^{(n)}$, which is in $\mathbf{V}$, and let μ_n denote the natural projection of $T^{(n+1)}$ onto $T^{(n)}$. Let also $\psi_n : S \to T^{(n)}$ be the target tupling of the morphisms φ_i, $i \leq n$, let T be the inverse limit $\lim_n T^{(n)}$ determined by the morphisms μ_n, and let $\pi_n : T \to T^{(n)}$ be the corresponding projection morphisms.

Note that each ψ_n is n-injective, and equals the composite of ψ_{n+1} and μ_n. Thus, there exists a (unique) morphism $\psi : S \to T$ such that the composite of ψ and π_n is ψ_n for each n. It follows from the n-injectivity of each ψ_n, that ψ is injective. Thus, S embeds in the inverse limit of an ω-diagram of preclones in $\mathbf{V}$, and we conclude by Lemma 4.20. $\square$

We note the following easy corollary of Proposition 4.21.

Corollary 4.22 *Let $\mathbf{V}$ be a pseudovariety of preclones. Let S be a finitary preclone such that distinct elements of equal rank can be separated by a morphism from S to a preclone in $\mathbf{V}$. Then $S \in \mathbf{V}$.*

Proof. For any distinct elements f, g of equal rank n, let $\varphi_{f,g} : S \to S_{f,g}$ be a morphism such that $S_{f,g} \in \mathbf{V}$ and $\varphi_{f,g}(f) \neq \varphi_{f,g}(g)$. For any integer n, let φ_n be the target tupling of the finite collection of morphisms $\varphi_{f,g}$ with $f, g \in S_n$. Then φ_n is injective on S_n and we conclude by Proposition 4.21. $\square$

4.4 Pseudovarieties of pg-pairs

The formal treatment pseudovarieties of *pg*-pairs is similar to the above treatment of pseudovarieties of preclones – but for the following remarks.

We define a *pseudovariety of pg-pairs* to be a class of finitary *pg*-pairs closed under finite direct product, sub-*pg*-pairs, quotients and finitary inverse limits of ω-diagrams. Our first remark is that, in this case, we do not need to mention finitary unions of ω-chains: indeed, finitary *pg*-pairs are finitely generated, so the union of an ω-chain, if it is finitary, amounts to a finite union.

Next, the notion of inverse limit of ω-diagrams of *pg*-pairs needs some clarification. Consider a sequence of morphisms of *pg*-pairs, say $\varphi_n : (S^{(n+1)}, A^{(n+1)}) \to (S^{(n)}, A^{(n)})$. That is, each φ_n is a preclone morphism from $S^{(n+1)}$ to $S^{(n)}$, which maps $A^{(n+1)}$ into $A^{(n)}$. We can then form the inverse limit $\lim_n S^{(n)}$ of the ω-diagram determined by the preclone morphisms φ_n, and the inverse limit $\lim_n A^{(n)}$ determined by the set mappings φ_n. The inverse limit $\lim_n (S^{(n)}, A^{(n)})$ of the ω-diagram determined by the morphisms of *pg*-pairs φ_n (as determined by the appropriate universal limit, see Remark 4.2) is the *pg*-pair (S, A), where $A = \lim_n A^{(n)}$ and S is the subpreclone of $\lim_n S^{(n)}$ generated by A. Recall

that this inverse limit is called finitary exactly when S is finitary and A is finite (see Example 4.3).

We now establish the close connection between this inverse limit and the inverse limit of the underlying ω-diagram of preclones, when the latter is finitary.

Proposition 4.23 *Let $\varphi_n\colon (S^{(n+1)}, A^{(n+1)}) \to (S^{(n)}, A^{(n)})$ be an ω-diagram of pg-pairs. Let $S = \lim_n S^{(n)}$ and let and $(T, A) = \lim_n (S^{(n)}, A^{(n)})$. If S is finitary, then $S = T$.*

Proof. We need to show that A generates S. Without loss of generality, we may assume that each φ_n maps $A^{(n+1)}$ surjectively onto $A^{(n)}$, and we denote by χ_n the restriction of φ_n to $A^{(n+1)}$. By definition, A is the inverse limit of the ω-diagram given by the χ_n, and we denote by $\rho_n\colon A \to A^{(n)}$ the corresponding projection. We also denote by χ_n and ρ_n the extensions of these mappings to preclone morphisms $A^{(n+1)}M \to A^{(n)}M$ and $AM \to A^{(n)}M$. It is not difficult to verify that AM is the inverse limit of the ω-diagram given by the χ_n, and that the ρ_n are the corresponding projections.

$$\begin{tikzcd}
 & A \arrow[d, "\rho_{n+1}"'] \arrow[dr, "\rho_n"] & \\
\cdots \arrow[r] & A^{(n+1)} \arrow[r, "\chi_n"'] & A^{(n)} \cdots
\end{tikzcd}
\qquad
\begin{tikzcd}
 & AM \arrow[d, "\rho_{n+1}"'] \arrow[dr, "\rho_n"] & \\
\cdots \arrow[r] & A^{(n+1)}M \arrow[r, "\chi_n"'] & A^{(n)}M \cdots
\end{tikzcd}$$

Moreover, each ρ_k is onto (even from A to $A^{(k)}$). Let indeed $a_k \in A^{(k)}$. Since the χ_n are onto, we can define by induction a sequence $(a_n)_{n\geq k}$ such that $\chi_n(a_{n+1}) = a_n$ for each $n \geq k$. This sequence can be completed with the iterated images of a_k by $\chi_{k-1}, \ldots, \chi_0$ to yield an element of A whose k-th projection is a_k.

Since $A^{(n)}$ generates $S^{(n)}$, the morphism $\psi_n\colon A^{(n)}M \to S^{(n)}$ induced by $\mathsf{id}_{A^{(n)}}$ is surjective. Moreover, the composites $\varphi_n \circ \psi_{n+1}$ and $\psi_n \circ \chi_n$ coincide.

$$\begin{tikzcd}
AM \arrow[dr, "\rho_{n+1}"'] \arrow[drr, "\rho_n"] \arrow[ddd, dashed, "\tau"'] & & \\
 & A^{(n+1)}M \arrow[r, "\chi_n"'] \arrow[d, "\psi_{n+1}"'] & A^{(n)}M \arrow[d, "\psi_n"] \\
 & S^{(n+1)} \arrow[r, "\varphi_n"] & S^{(n)} \\
S \arrow[ur, "\pi_{n+1}"] \arrow[urr, "\pi_n"'] & &
\end{tikzcd}$$

It follows that the morphisms $\psi_n \circ \rho_n \colon AM \to S^{(n)}$ and $\varphi_n \circ \psi_{n+1} \circ \rho_{n+1}$ coincide, and hence there exists a morphism $\tau\colon AM \to S$ such that $\pi_n \circ \tau = \psi_n \circ \rho_n$ for each n. Since ρ_n and ψ_n are onto, it follows that each π_n is surjective.

We now use the fact that S is finitary. By Lemma 4.14, π_n is k-injective for each large enough n. Let now $s \in S_k$. We want to show that $s \in \tau(AM)$. Let n_k be such that π_n is k-injective for each $n \geq n_k$. We can choose an element $t_{n_k} \in A^{(n_k)}M$ such that $\psi_{n_k}(t_{n_k}) = \pi_{n_k}(s)$. Then, by induction, we can construct a sequence $(t_n)_n$ of elements such that $\chi_n(t_{n+1}) = t_n$ for each $n \geq 0$. We need to show that $\psi_n(t_n) = \pi_n(s)$ for each n.

This equality is immediate for $n \leq n_k$, and we assume by induction that it holds for some $n \geq n_k$. We have

$$\varphi_n(\psi_{n+1}(t_{n+1})) = \psi_n(\chi_n(t_{n+1})) = \psi_n(t_n) = \pi_n(s) = \varphi_n(\pi_{n+1}(t_{n+1})).$$

Since π_n and π_{n+1} are surjective, since they are injective on S_k, and since $\varphi_n \circ \pi_{n+1} = \pi_n$, we find that φ_n is injective on $S_k^{(n+1)}$, and hence $\psi_{n+1}(t_{n+1}) = \pi_{n+1}(s)$, as expected.

Thus $(t_n)_n \in AM$ and $\tau(t) = s$, which concludes the proof that S is generated by A. □

5 Varieties of tree languages

Let $\mathcal{V} = (\mathcal{V}_{\Sigma,k})_{\Sigma,k}$ be a collection of nonempty classes of recognizable tree languages $L \subseteq \Sigma M_k$, where Σ runs over the finite ranked alphabet and k runs over the nonnegative integers. We call $\mathcal{V}$ a *variety of tree languages*, or a *tree language variety*, if each $\mathcal{V}_{\Sigma,k}$ is closed under the Boolean operations, and $\mathcal{V}$ is closed under inverse morphisms between free preclones generated by finite ranked sets, and under quotients defined as follows. Let $L \subseteq \Sigma M_k$ be a tree language, let k_1 and k_2 be nonnegative integers, $u \in \Sigma M_{k_1+1+k_2}$ and $v \in \Sigma M_{n,k}$. Then the *left quotient* $(u, k_1, k_2)^{-1}L$ and the *right quotient* Lv^{-1} are defined by

$$\begin{aligned} (u, k_1, k_2)^{-1}L &= \{t \in \Sigma M_n \mid u \cdot (\mathbf{k}_1 \oplus t \oplus \mathbf{k}_2) \in L\} \quad \text{where } k = k_1 + n + k_2 \\ Lv^{-1} &= \{t \in \Sigma M_n \mid t \cdot v \in L\}, \end{aligned}$$

that is, $(u, k_1, k_2)^{-1}L$ is the set of elements of ΣM_n for which $(u, k_1, \mathbf{n}, k_2)$ is an L-context, and Lv^{-1} is the set of elements of ΣM_n for which $(\mathbf{1}, 0, v, 0)$ is an L-context. Below we will write just $u^{-1}L$ for $(u, k_1, k_2)^{-1}L$ if k_1 and k_2 are understood, or play no role.

A *literal variety of tree languages* is defined similarly, but instead of closure under inverse morphisms between finitely generated free preclones, we require closure under inverse morphisms between finitely generated free *pg*-pairs. Thus, if $L \subseteq \Sigma M_k$ is in a literal variety $\mathcal{V}$ and $\varphi \colon \Delta M \to \Sigma M$ is a preclone morphism with Σ, Δ finite and $\varphi(\Delta) \subseteq \Sigma$, then $\varphi^{-1}(L)$ is also in $\mathcal{V}$.

5.1 Varieties of tree languages vs. pseudovarieties of preclones

The aim of this section is to prove an Eilenberg correspondence between pseudovarieties of preclones (resp. *pg*-pairs), and varieties (resp. literal varieties) of tree languages. For each pseudovariety $\mathbf{V}$ of preclones (resp. *pg*-pairs), let $\mathsf{var}(\mathbf{V}) = (\mathcal{V}_{\Sigma,k})_{\Sigma,k}$, where $\mathcal{V}_{\Sigma,k}$ denotes the class of the tree languages $L \subseteq \Sigma M_k$ whose syntactic preclone (resp. *pg*-pair) belongs to $\mathbf{V}$. It follows from Proposition 3.2 that $\mathsf{var}(\mathbf{V})$ consists of all those tree languages that can be recognized by a preclone (resp. *pg*-pair) in $\mathbf{V}$.

Conversely, if $\mathcal{W}$ is a variety (resp. a literal variety) of tree languages, we let $\mathsf{psv}(\mathcal{W})$ be the class of all finitary preclones (resp. *pg*-pairs) that only accept languages in $\mathcal{W}$, *i.e.*, $\alpha^{-1}(F) \subseteq \Sigma M_k$ belongs to $\mathcal{W}$, for all morphisms $\alpha \colon \Sigma M \to S$ (resp. $\alpha \colon (\Sigma M, \Sigma) \to (S, A)$), $k \geq 0$ and $F \subseteq S_k$.

Theorem 5.1 *The mappings* var *and* psv *are mutually inverse lattice isomorphisms between the lattice of pseudovarieties of preclones (resp. pg-pairs) and the lattice of varieties (resp. literal varieties) of tree languages.*

Proof. We only prove the theorem for pseudovarieties of *pg*-pairs and literal varieties of tree languages. It is clear that for each pseudovariety $\mathbf{V}$ of finitary *pg*-pairs, if $\mathsf{var}(\mathbf{V}) = (\mathcal{V}_{\Sigma,k})_{\Sigma,k}$, then each $\mathcal{V}_{\Sigma,k}$ is closed under complementation and contains the languages $\emptyset$ and ΣM_k. The closure of $\mathcal{V}_{\Sigma,k}$ under union follows in the standard way from the closure of $\mathbf{V}$ under direct product: if $L, L' \subseteq \Sigma M_k$ are recognized by morphisms into *pg*-pairs (S, A) and (S', A') in $\mathbf{V}$, then $L \cup L'$ is recognized by a morphism into $(S, A) \times (S', A')$. Thus $\mathcal{V}_{\Sigma,k}$ is closed under the Boolean operations.

We now show that $\mathcal{V}$ is closed under quotients. Let $L \subseteq \Sigma M_k$ be in $\mathcal{V}_{\Sigma,k}$, let $\alpha \colon (\Sigma M, \Sigma) \to (S, A)$ be a morphism recognizing L with $(S, A) \in \mathbf{V}$ and $L = \alpha^{-1}\alpha(L)$, and let $F = \alpha(L)$. Let (u, k_1, v, k_2) be an n-ary context, that is, $u \in \Sigma M_{k_1+1+k_2}$, $v \in \Sigma M_{n,\ell}$ and $k_1 + \ell + k_2 = k$. Now let $F' = \{f \in S_\ell \mid \alpha(u) \cdot (\mathbf{k}_1 \oplus f \oplus \mathbf{k}_2) \in F\}$. Then for any $t \in \Sigma M_\ell$, $\alpha(t) \in F'$ if and only if $\alpha(u) \cdot (\mathbf{k}_1 \oplus \alpha(t) \oplus \mathbf{k}_2) \in F$, if and only if $\alpha(u \cdot (\mathbf{k}_1 \oplus t \oplus \mathbf{k}_2)) \in F$ iff $u \cdot (\mathbf{k}_1 \oplus t \oplus \mathbf{k}_2) \in L$. Thus, $\alpha^{-1}(F') = (u, k_1, k_2)^{-1}L$, which is therefore in $\mathcal{V}_{\Sigma,\ell}$. Now let $F'' = \{f \in M_n : f \cdot \alpha(v) \in L\}$. It follows as above that $Lv^{-1} = \alpha^{-1}(F'')$ and hence $Lv^{-1} \in \mathcal{V}_{\Sigma,n}$.

Before we proceed, let us observe that we just showed the following: if $L \subseteq \Sigma M_k$ is a recognizable tree language, then for each $n \geq 0$ there are only finitely many distinct sets of the form $((u, k_1, k_2)^{-1}L)v^{-1}$, where (u, k_1, v, k_2) is an n-ary context of ΣM_k.

Next, let $\varphi\colon (\Sigma M, \Sigma) \to (\Delta M, \Delta)$ be a morphism of *pg*-pairs and $L \subseteq \Delta M_k$. If L is recognized by a morphism $\alpha\colon (\Delta M, \Delta) \to (S, A)$, then $\varphi^{-1}(L)$ is recognized by the composite morphism $\varphi \circ \alpha$, and the closure of $\mathcal{V}$ by inverse morphisms between free *pg*-pairs follows immediately. Thus the mapping var does associate with each pseudovariety of *pg*-pairs a literal variety of tree languages, and it clearly preserves the inclusion order.

Now consider the mapping psv: we first verify that if $\mathcal{W}$ is a literal variety of tree languages, then the class $\mathsf{psv}(\mathcal{W})$ is a pseudovariety. Recall that, if $(S, A) < (T, B)$, then any language recognized by (S, A) is also recognized by (T, B), so if each language recognized by (T, B) belongs to $\mathcal{W}$, then the same holds for (S, A). Note also that any language recognized by the direct product $(S, A) \times (T, B)$ is a finite union of intersections of the form $L \cap M$, where L is recognized by (S, A) and M by (T, B); thus $\mathsf{psv}(\mathcal{W})$ is closed under binary direct products. Finally, if $(S, A) = \lim_n(S^{(n)}, A^{(n)})$ is the finitary inverse limit of an ω-diagram of finitary *pg*-pairs, then Lemma 4.14 shows that the languages recognized by (S, A) are recognized by almost all of the $(S^{(n)}, A^{(n)})$. Thus $(S, A) \in \mathsf{psv}(\mathcal{W})$, which concludes the proof that $\mathsf{psv}(\mathcal{W})$ is a pseudovariety of *pg*-pairs.

Let $\mathcal{W}$ be a literal variety of tree languages, and let $\mathcal{V} = \mathsf{var}(\mathsf{psv}(\mathcal{W}))$. We now show that $\mathcal{V} = \mathcal{W}$. Since $\mathcal{V}$ consists of all the tree languages recognized by a *pg*-pair in $\mathbf{W} = \mathsf{psv}(\mathcal{W})$, it is clear that $\mathcal{V} \subseteq \mathcal{W}$. Now let $L \in \mathcal{W}_{\Sigma,k}$, and let (M_L, A_L) be its syntactic *pg*-pair. To prove that $(M_L, A_L) \in \mathbf{W}$, it suffices to show that if $\alpha\colon (\Sigma M, \Sigma) \to (M_L, A_L)$ is a morphism of *pg*-pairs and $x \in M_L$, then $\alpha^{-1}(x) \in \mathcal{W}$. Since a morphism of *pg*-pairs maps generators to generators, up to renaming and identifying letters (which can be done by morphisms between free *pg*-pairs), we may assume that α is the syntactic morphism of L. Thus $\alpha^{-1}(x)$ is an equivalence class $[w]$ in the syntactic congruence of L, and hence

$$\begin{aligned}\alpha^{-1}(x) = & \bigcap_{w \in ((u,k_1,k_2)^{-1}L)v^{-1}} ((u, k_1, k_2)^{-1}L)v^{-1} \\ & \cap \bigcap_{w \notin ((u,k_1,k_2)^{-1}L)v^{-1}} ((u, k_1, k_2)^{-1}\overline{L})v^{-1}\end{aligned}$$

where $\overline{L}$ denotes the complement of L. If x has rank n, the intersections in this formula run over n-ary contexts (u, k_1, v, k_2), and as observed above, these intersections are in fact finite. It follows that that $\alpha^{-1}(x) \in \mathcal{V}$. This concludes the verification that $\mathcal{V} = \mathcal{W}$, so $\mathsf{var} \circ \mathsf{psv}$ is the identity mapping, and in

particular var is surjective and psv is injective.

It is clear that both maps var and psv preserve the inclusion order. In order to conclude that they are mutually reciprocal bijections, it suffices to verify that var is injective. If $\mathbf{V}$ and $\mathbf{W}$ are pseudovarieties such that $\mathsf{var}(\mathbf{V}) = \mathsf{var}(\mathbf{W}) = \mathcal{V}$, then a tree language is in $\mathcal{V}$ if and only if its syntactic preclone is in $\mathbf{V}$, if and only if its syntactic preclone is in $\mathbf{W}$. Thus $\mathbf{V}$ and $\mathbf{W}$ contain the same syntactic preclones, and it follows from Corollary 4.8 that $\mathbf{V} = \mathbf{W}$. $\square$

Remark 5.2 Three further variety theorems for finite trees exist in the literature. They differ from the variety theorem proved above since they use a different notion of morphism, quotient, and syntactic algebra. The variety theorem in [1,35] is formulated for tree language varieties over some fixed ranked alphabet and the morphisms are homomorphisms between finitely generated free algebras, whereas the "general variety theorem" of [36] allows for tree languages over different ranked alphabets and a more general notion of morphism, closely related to the morphisms of free pg-pairs. On the other hand, the morphisms in [19] are much more general than those in either [1,35,36] or the present paper, they even include nonlinear tree morphisms that allow for the duplication of a variable. Another difference is that the tree language varieties in [1,35,36] involve only left quotients, whereas the one presented here (and the varieties of [19]) are defined using two sided quotients. The notion of syntactic algebra is also different in these papers: minimal tree automata in [1,35], a variant of minimal tree automata in [36], minimal clone (or Lawvere theory) in [19], and minimal preclone, or pg-pair, here. We refer to [19, Section 14] for a more detailed comparative discussion.

As noted above, the abundance of variety theorems for finite trees is due to the fact that there are several reasonable ways of defining morphisms and quotients, and a choice of these notions is reflected by the corresponding notion of syntactic algebra. No variety theorem is known for the 3-sorted algebras proposed in [41]. $\square$

5.2 Examples of varieties of tree languages

5.2.1 Small examples

As a practice example, we describe the variety of tree languages associated with the pseudovariety $\langle T_\exists \rangle$ generated by $T_\exists$ (see Section 2.3.2).

Let Σ be a finite ranked alphabet and let $L \subseteq \Sigma M_k$ be a tree language accepted by a preclone in $\langle T_\exists \rangle$. Then the syntactic preclone S of L lies in $\langle T_\exists \rangle$. Recall that a syntactic preclone is finitely generated and finitely determined: it follows

from Proposition 4.16 that S divides a product of a finite number of copies of $T_\exists$. By a standard argument, L is therefore a (positive) Boolean combination of languages recognized by a morphism from ΣM to $T_\exists$.

Now let $\tau\colon \Sigma M \to T_\exists$ be a morphism. As discussed in Section 3.3, a tree language in ΣM recognized by τ is either of the form $K_k(\Sigma')$ for some $\Sigma' \subseteq \Sigma$, or it is the complement of such a language. From there, and using the same reasoning as in the analogous case concerning word languages, one can verify that a language $L \in \Sigma M_k$ is accepted by a preclone in $\langle T_\exists \rangle$ if and only if L is a Boolean combination of languages of the form $K_k(\Sigma')$ ($\Sigma' \subseteq \Sigma$), or equivalently, L is a Boolean combination of languages of the form $L_k(\Sigma')$, $\Sigma' \subseteq \Sigma$, where $L_k(\Sigma)$ is the set of all Σ-trees of rank k, for which the set of node labels is exactly Σ'.

Similarly – and referring again to Section 3.3 for notation – one can give a description of the variety of tree languages associated with the pseudovariety $\langle T_p \rangle$, or the pseudovariety $\langle T_{p,q} \rangle$, using the languages of the form $K_k(\exists_p^r)$ or $K_k(\exists_{p,q}^r)$ instead of the $K_k(\exists)$.

5.2.2 *FO*[**Succ**]-*definable tree languages*

In a recent paper [3], Benedikt and Ségoufin considered the class of $FO[\mathbf{Succ}]$-definable tree languages. Note that the logical language used in $FO[\mathbf{Succ}]$ does not allow the predicate $<$, and $FO[\mathbf{Succ}]$ is a fragment of $FO[<]$. We refer the reader to [3] for precise definitions, and we point out here that the characterization established there can be expressed in the framework developed in the present paper.

More precisely, the results of Benedikt and Ségoufin establish that $FO[\mathbf{Succ}]$-definable tree languages form a variety of languages, and that the corresponding pseudovariety of preclones consists of the preclones S such that

(1) the semigroup S_1 satisfies $x^\ell = x^{\ell+1}$ and $exfyezf = ezfyexf$ for all elements e, f, x, y, z such that $e = e^2$ and $f = f^2$ and for $\ell = |S_1|$;
(2) for each $x \in S_2$, $e \in S_1$ such that $e = e^2$, and $s, t \in S_0$, $x \cdot (e \cdot s \oplus e \cdot t) = x \cdot (e \cdot t \oplus e \cdot s)$.

In particular, $FO[\mathbf{Succ}]$-definability is decidable for regular tree languages.

It is clearly argued in [3] that $FO[\mathbf{Succ}]$-definable tree languages are exactly the locally threshold testable languages, for general model-theoretic reasons, but that this fact alone does not directly yield a decision procedure. The result stated above is analogous to the characterization of $FO[\mathbf{Succ}]$-definability for recognizable word languages - more precisely, Condition (1) suffices for languages of words and their syntactic semigroups. Condition (2), which makes

sense in trees but not in words, must be added to the other one to characterize $FO[\mathbf{Succ}]$-definability for tree languages.

5.2.3 Some classes of languages definable in modal logic

Bojańczyk and Walukiewicz also characterized interesting logically defined classes of tree languages [5]. Again, their results are not couched in terms of preclones, but they can conveniently be expressed in this way.

These authors consider three fragments of CTL*: $TL(\mathsf{EX})$, $TL(\mathsf{EF})$ and $TL(\mathsf{EX}+\mathsf{EF})$. Here EX (resp. EF) denotes the modality whereby a tree t satisfies $\mathsf{EX}\varphi$ (resp. $\mathsf{EF}\varphi$) if some child of the root (resp. some node properly below the root) of t satisfies φ. The set of formulas constructed using one or both of these modalities, plus Boolean operations and letter constants form the logical languages $TL(\mathsf{EX})$, $TL(\mathsf{EF})$ and $TL(\mathsf{EX}+\mathsf{EF})$.

Bojańczyk and Walukiewicz first observe that a tree language L is $TL(\mathsf{EX})$-definable if and only if there exists an integer k such that membership of a tree t in L depends only on the fragment of t consisting of the nodes of depth at most k. They then show that these tree languages form a variety, and the corresponding pseudovariety of preclones consists of the preclones S such that the semigroup S_1 satisfies $ex = e$ for each idempotent e. Note that this is exactly the same characterization as for languages of finite words [31].

For the characterization of $TL(\mathsf{EF})$-definable languages, let us first define the following relation on a preclone S : if $s, t \in S_n$, we say that $s \preceq t$ if $s = u \cdot t$ for some $u \in S_1$. It is easily verified that $\preceq$ is a quasi-order. (The direction of the order is reversed from that used by Bojańczyk and Walukiewicz, to enhance the analogy with the $\mathcal{R}$- and $\mathcal{L}$-orders in semigroup theory).

Now let (S, A) be the syntactic pg-pair of a tree language $L \subseteq \Sigma M_0$. Then L is $TL(\mathsf{EF})$-definable if and only if

- S_1 satisfies the pseudo-identity $v(uv)^\omega = (uv)^\omega$ (where x^ω designates the unique idempotent which is a power of x); that is, S_1 is $\mathcal{L}$-trivial, and equivalently, the relation $\preceq$ is an order relation;
- $a \cdot (s_1 \oplus \cdots \oplus s_n) = a \cdot (s_{\pi(1)} \oplus \cdots \oplus s_{\pi(n)})$ for each $a \in A_n$ and $s_1, \ldots, s_n \in S_0$, and for each permutation π of $[n]$;
- $a \cdot (s_1 \oplus s_2 \oplus s_3 \oplus \cdots \oplus s_n) = a \cdot (s_2 \oplus s_2 \oplus s_3 \oplus \cdots \oplus s_n)$ for each $a \in A_n$ and $s_1, \ldots, s_n \in S_0$ such that $s_2 \preceq s_1$;
- if $b, c \in A_p$ and $y \in S_{p,0}$ are such that, for each $d \in A_p$, we have $d \cdot (b \cdot y \oplus \cdots \oplus b \cdot y) = d \cdot y = d \cdot (c \cdot y \oplus \cdots \oplus c \cdot y)$, then $a \cdot (z \oplus b \cdot y) = a \cdot (z \oplus c \cdot y)$ for each $a \in A_n$ and $z \in S_{n-1,0}$.

This characterization directly implies the decidability of $TL(\mathsf{EF})$-definability.

Bojańczyk and Walukiewicz also give an interesting characterization of the $TL(\mathsf{EF})$-definable languages in terms of so-called *type dependency*. In particular, they show that a tree language is $TL(\mathsf{EF})$-definable if and only if its syntactic preclone S is such that, whenever a is the syntactic equivalence class of a letter in Σ_n, and the t_i's are syntactic equivalence classes of trees in ΣM_0, then the value of a product $a \cdot (t_1 \oplus \cdots \oplus t_n)$ depends only on a and on the set $\{t \mid t_i \preceq t \text{ for some } 1 \leq i \leq n\}$.

The characterization of $TL(\mathsf{EX} + \mathsf{EF})$-definable languages given in [5] can also be restated in similar – albeit more complex – terms.

5.2.4 $FO[<]$-definable tree languages

The characterization and decidability of $FO[<]$-definable regular tree languages is an open problem that has attracted some efforts along the years, as discussed in the introduction.

We obtained an algebraic characterization of $FO[<]$-definable regular tree languages in terms of pseudovarieties of preclones, as is reported in [20]. A detailed report of this result will appear in [21], and the present paper lays the foundations for this proof.

Let us note here that this characterization is analogous to the characterization of $FO[<]$-definable languages of finite words in the following sense: it is established in [21] that $FO[<]$-definable tree languages form a variety of tree languages, whose associated pseudovariety of preclones is the least pseudovariety containing the preclone $T_\exists$ and closed under a suitable notion of block product. It was pointed out in Example 2.5 that the rank 1 elements of $T_\exists$ form the 2-element monoid $U_1 = \{1, 0\}$, and it is a classical result of language theory that the least pseudovariety of monoids containing U_1 and closed under block product is associated with the variety of $FO[<]$-definable word languages [37].

It is also known that, in the word case, this pseudovariety is exactly that of aperiodic monoids, and membership in it is decidable, which shows that $FO[<]$-definability is decidable for recognizable word languages. At the moment, we do not have an analogue of this result, and we do not know whether $FO[<]$-definability is decidable for regular tree languages.

Our result [20,21] actually applies to a larger class of logically defined regular tree languages, based on the use of Lindström quantifiers. First-order logic is thus a particular case of our result, which also yields (for instance) an algebraic characterization of first-order logic with modular quantifiers added.

References

[1] J. Almeida, On pseudovarieties, varieties of languages, filters of congruences, pseudoidentities and related topics, *Algebra Universalis*, 27 (1990), 333–350.

[2] A. Arnold, M. Dauchet, Théorie des magmoïdes. I. and II. (in French), *RAIRO Theoretical Informatics and Applications*, 12 (1978), 235–257, 3 (1979), 135–154.

[3] M. Benedikt, L. Ségoufin, Regular tree languages definable in *FO*. In: *STACS 2005* (V. Diekert, B. Durand eds.), Lect. Notes in Computer Science, Springer, to appear.

[4] S. L. Bloom, Z. Ésik, *Iteration Theories*, Springer, 1993.

[5] M. Bojańczyk, I. Walukiewicz, Characterizing EF and EX tree logics, in *CONCUR 2004* (P. Gardner, N. Yoshida, eds.), Lect. Notes in Computer Science 3170, Springer, 2004.

[6] J. R. Büchi, Weak second-order arithmetic and finite automata, *Z. Math. Logik Grundlagen Math.*, 6 (1960), 66–92.

[7] J. Cohen, J.-E. Pin, D. Perrin, On the expressive power of temporal logic, *J. Computer and System Sciences*, 46 (1993), 271–294.

[8] H. Comon, M. Dauchet, R. Gilleron, F. Jacquemard, D. Lugiez, S. Tison, M. Tommasi, *Tree Automata Techniques and Applications*, available on: `http://www.grappa.univ-lille3.fr/tata` (release October 2002).

[9] B. Courcelle, The monadic second-order logic of graphs. I. Recognizable sets of finite graphs, *Information and Computation*, 85 (1990), 12–75.

[10] B. Courcelle, Basic notions of universal algebra for language theory and graph grammars, *Theoret. Computer Science*, 163 (1996), 1–54.

[11] B. Courcelle, The expression of graph properties and graph transformations in monadic second order logic. In: G. Rozenberg (ed.) *Handbook of Graph Grammars and Computing by Graph Transformations*, vol. 1, World Scientific, 1997, 313–400.

[12] B. Courcelle, P. Weil, The recognizability of sets of graphs is a robust property, to appear.

[13] K. Denecke, S. L. Wismath, *Universal Algebra and Applications in Theoretical Computer Science*, Chapman and Hall, 2002.

[14] V. Diekert, *Combinatorics on Traces*, Lect. Notes in Computer Science 454, Springer, 1987.

[15] J. Doner, Tree acceptors and some of their applications, *J. Comput. System Sci.*, 4 (1970), 406–451.

[16] S. Eilenberg, *Automata, Languages, and Machines*, vol. A and B, Academic Press, 1974 and 1976.

[17] S. Eilenberg, J. B. Wright, Automata in general algebras, *Information and Control*, 11 (1967), 452–470.

[18] C. C. Elgot, Decision problems of finite automata design and related arithmetics, *Trans. Amer. Math. Soc.*, 98 (1961), 21–51.

[19] Z. Ésik, A variety theorem for trees and theories, *Publicationes Mathematicae*, 54 (1999), 711–762.

[20] Z. Ésik, P. Weil, On logically defined recognizable tree languages. In: *Proc. FST TCS 2003* (P. K. Pandya, J. Radhakrishnan eds.), Lect. Notes in Computer Science 2914 (2003), Springer, 195-207.

[21] Z. Ésik, P. Weil, Algebraic characterization of logically defined tree languages, in preparation.

[22] D. M. Gabbay, A. Pnueli, S. Shelah, J. Stavi, On the temporal analysis of fairness. In: proc. *12th ACM Symp. Principles of Programming Languages*, Las Vegas, 1980, 163–173.

[23] F. Gécseg, M. Steiby, *Tree Automata*, Akadémiai Kiadó, Budapest, 1984.

[24] F. Gécseg, M. Steiby, Tree languages. In: G. Rozenberg, A. Salomaa, eds. *Handbook of Formal Languages*, vol. 3, Springer, 1997, 1-68.

[25] G. Grätzer, *Universal Algebra*, Springer, 1979.

[26] U. Heuter, First-order properties of trees, star-free expressions, and aperiodicity. In: *STACS 88* (eds. R. Cori, M. Wirsing), Lect. Notes in Computer Science 294, Springer, 1988, 136–148.

[27] J. A. Kamp, Tense logic and the theory of linear order, Ph. D. Thesis, UCLA, 1968.

[28] S. MacLane, *Categories for the Working Mathematician*, Springer, 1971.

[29] R. McNaughton, S. Papert, *Counter-Free Automata*, MIT Press, 1971.

[30] J. Mezei, J. B. Wright, Algebraic automata and context-free sets, *Information and Control*, 11 (1967), 3–29.

[31] J.-E. Pin, *Variétés de langages formels*, Masson, Paris (1984); English translation: *Varieties of formal languages*, Plenum, New-York (1986).

[32] A. Potthoff, Modulo counting quantifiers over finite trees. In: *CAAP '92* (J.-C. Raoult ed.), Lect. Notes in Computer Science 581, Springer, 1992, 265–278.

[33] A. Potthoff, First order logic on finite trees. In: *TAPSOFT '95* (P.D. Mosses, M. Nielsen, M.I. Schwartzbach eds.), Lect. Notes in Computer Science 915, Springer, 1995, 125–139.

[34] M. P. Schützenberger, On finite monoids having only trivial subgroups. *Information and Control*, 8 (1965), 190–194.

[35] M. Steinby, A theory of tree language varieties. In: *Tree automata and languages* (M. Nivat, A. Podelski, eds.), North Holland, 1992, 57–81.

[36] M. Steinby, General varieties of tree languages. *Theoret. Comput. Sci.*, 205 (1998), 1–43.

[37] H. Straubing, *Finite Automata, Formal Logic, and Circuit Complexity,* Birkhaüser, Boston, MA, 1994.

[38] J. W. Thatcher, J. B. Wright, Generalized finite automata theory with an application to a decision problem of second-order logic, Math. Systems Theory, 2 (1968), 57–81.

[39] W. Wechler, *Universal algebra*, EATCS Monographs on Theoretical Computer Science 10, Springer, 1992.

[40] P. Weil, Algebraic recognizability of languages. In: *MFCS 2004* (J. Fiala, V. Koubek, J. Kratochvíl eds.), Lect. Notes in Computer Science 3153 (Springer, 2004) 149–175.

[41] Th. Wilke, An algebraic characterization of frontier testable tree languages, *Theoret. Comput. Sci.*, 154 (1996), 85–106.